

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 16:48:09 ; Search time 53 Seconds
(without alignments)
4306.581 Million cell updates/sec

Title: US-10-006-091-1
Perfect score: 7691
Sequence: 1 ATRYYLGAVELSWDYMQSD.....WVHQIALRMEVLGCEAQQDLY 1438

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7691	100.0	1438	21 AAB01262	B-domain deleted f
2	7691	100.0	1457	20 AAY21675	Beta-domain delete
3	7684	99.9	1457	19 AAW46246	Human factor VIII
4	7684	99.9	1457	19 AAW44372	Human factor VIII
5	7674	99.8	1471	18 AAW23414	Human B-domain del
6	7674	99.8	1471	22 AAB67959	Amino acid sequenc
7	7659.5	99.6	1440	12 AAR12971	Factor VIII:SQ. U
8	7626	99.2	1459	22 AAE10832	Human factor VIII
9	7616	99.0	1459	22 AAE10833	Human factor VIII

10	7613	99.0	1459	22 AAE10827	Human factor VIII
11	7611.5	99.0	1516	9 AAP80265	Modified factor VII
12	7592	98.7	1424	22 AAB48842	Mutant mature huma
13	7592	98.7	1424	23 AAO18622	Human mature B-dom
14	7588	98.7	1447	23 ABG92541	SArg B-domain-dele
15	7581	98.6	1425	9 AAP80267	Modified factor VI
16	7578	98.5	1424	9 AAP80268	Modified factor VI
17	7578	98.5	1445	23 ABG92540	LE B-domain-delete
18	7574	98.5	1424	10 AAP91169	Sequence of 740 Ar
19	7434	96.7	1661	18 AAW18670	Factor VIII-dB695-
20	7353.5	95.6	1383	18 AAW33227	Procoagulant-activ
21	7345.5	95.5	1383	18 AAW33229	Procoagulant-activ
22	7342.5	95.5	1383	18 AAW33228	Procoagulant-activ
23	7234	94.1	2332	14 AAW43457	Human factor VIII.
24	7234	94.1	2332	19 AAW53483	Human factor VIII.
25	7234	94.1	2332	20 AAY31594	Human factor VIII
26	7234	94.1	2332	22 AAE10826	Human mature wild-
27	7234	94.1	2332	22 AAE11200	Human factor VIII
28	7234	94.1	2332	22 AAE50465	Human factor VIII.
29	7234	94.1	2332	23 AAU79869	Human factor VIII.
30	7234	94.1	2351	20 AAY21676	Factor VIII protei
31	7231.5	94.0	2342	18 AAW11422	Active Factor VIII
32	7231	94.0	2351	18 AAW10591	Factor VIII:C (Phe
33	7231	94.0	2351	18 AAW13496	Factor VIII:C (Arg
34	7230.5	94.0	2344	18 AAW11432	Active Factor VIII
35	7230.5	94.0	2344	18 AAW11410	Active Factor VIII
36	7230	94.0	2332	23 AAU79870	Human factor VIII
37	7230	94.0	2332	23 AAU79872	Human factor VIII
38	7230	94.0	2345	18 AAW11415	Active Factor VIII
39	7230	94.0	2345	18 AAW11403	Active Factor VIII
40	7229.5	94.0	2346	18 AAW11421	Active Factor VIII
41	7229.5	94.0	2346	18 AAW11431	Active Factor VIII
42	7229.5	94.0	2346	18 AAW11434	Active Factor VIII
43	7229	94.0	2347	18 AAW11411	Active Factor VIII
44	7229	94.0	2347	18 AAW11402	Active Factor VIII
45	7229	94.0	2351	18 AAW10592	Factor VIII:C (Tyr

ALIGNMENTS

RESULT 1
AAB01262 ID AAB01262 standard; protein; 1438 AA.

XX AAB01262;

XX 25-SEP-2000 (first entry)

XX B-domain deleted factor VIII sequence.

XX Factor VIII; procoagulant; adenovirus; adeno-associated strain;
XX gene therapy; human Burkitt's lymphoma; HKB; therapy;
XX therapeutic protein; vector; Epstein-Barr virus; human.

XX Homo sapiens.

XX WO200034505-A1.

XX 15-SEP-2000.

XX 08-DEC-1999; 99WO-US29169.

XX 10-DEC-1998; 98US-0209916.

XX (FARB) BAYER CORP.

XX Cho M, Chan SY, Kelsey W, Yee H;

XX WPI; 2000-431311/37.

XX Producing cells expressing a protein having factor VIII procoagulant
XX activity especially, human factor VIII in an industrial scale, involves

PT expressing a vector comprising a sequence coding for factor VIII in
PT human cells
XX
PS Claim 7; Fig 1; 27pp; English.
XX
CC Producing cells expressing a protein having factor VIII procoagulant
CC activity, comprises contacting the cells with a vector comprising a
CC selectable marker and a sequence coding for the protein having factor
CC VIII procoagulant activity operably linked to a promoter. The cells
CC are then selected and individual clones expressing high levels of the
CC protein are isolated from the selected cells. The cells produced by
CC the method are not only useful for producing protein having factor
CC VIII procoagulant activity but also for producing adenovirus and
CC adeno-associated virus strains for gene therapy. The advantage of
CC having cells producing protein with factor VIII procoagulant activity
CC is that factor VIII protein can be produced on an industrial scale
CC in the range of 2-4 pg/cell/day. Human Burkitt's lymphoma (HKB)
CC cells provide a protein-free production system to produce not only
CC B-domain deleted factor VIII but also other therapeutic proteins. The
CC vector used in the method preferably comprises B-domain deleted
CC factor VIII (BDD-FVIII), a transcriptional unit for BDD-FVIII and a
CC selectable marker, dihydrofolate reductase (dhfr). In addition, a
CC terminal repeat sequence from Epstein-Barr virus is inserted into the
CC vector to increase integration efficiency.
XX
SQ Sequence 1438 AA;
Query Match 100.0%; Score 7691; DB 21; Length 1438;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATRRYLGAVELSWDMQSDLGELPDARFPVRPKSPFNTSVYKKTFLVEFTVHLFN 60
DB 1 ATRRYLGAVELSWDMQSDLGELPDARFPVRPKSPFNTSVYKKTFLVEFTVHLFN 60
QY 61 IAKPRPPMGLLGPTIQAEBYDTVTITLKNASHPVSLHAGVSVKASEGAEYDDQTSQ 120
DB 61 IAKPRPPMGLLGPTIQAEBYDTVTITLKNASHPVSLHAGVSVKASEGAEYDDQTSQ 120
QY 121 REKEDKVPFGSGSHYVWVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR 180
DB 121 REKEDKVPFGSGSHYVWVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR 180
QY 181 EGS�AKEKTQTLHKFTLLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVANGYNR 240
DB 181 EGS�AKEKTQTLHKFTLLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVANGYNR 240
QY 241 SLPLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPTITFLTAQTLL 300
DB 241 SLPLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPTITFLTAQTLL 300
QY 301 MDLQGFLLFCHTSSHQHDMAYVKVDCPEEPQLRMKNNEAEYDDDLTDSMDVVRP 360
DB 301 MDLQGFLLFCHTSSHQHDMAYVKVDCPEEPQLRMKNNEAEYDDDLTDSMDVVRP 360
QY 361 DDNDSPSFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVLAPODRSRYKSQYLNNGPQRIG 420
DB 361 DDNDSPSFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVLAPODRSRYKSQYLNNGPQRIG 420
QY 421 RXYKXVRFMAYTDTFTKTRATOHESGILGPLLYGEVGDTLIIIPKQASRPYNIYPHGI 480
DB 421 RXYKXVRFMAYTDTFTKTRATOHESGILGPLLYGEVGDTLIIIPKQASRPYNIYPHGI 480
QY 481 TDVRLPYSLRLPKGVKHLKDFILPGEIPKYKMTVTVEDGPKSDPRCLTRYYSFVNME 540
DB 481 TDVRLPYSLRLPKGVKHLKDFILPGEIPKYKMTVTVEDGPKSDPRCLTRYYSFVNME 540
QY 541 RLASGLIGPLLI CYKESVDQGNQIMSDKRNVLIFSVDENKRSWLTENIORFLPNPAG 600
DB 541 RLASGLIGPLLI CYKESVDQGNQIMSDKRNVLIFSVDENKRSWLTENIORFLPNPAG 600
QY 601 VQLEDPFOASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 660
DB 601 VQLEDPFOASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 660

DB 601 VQLEDPFOASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 660
QY 661 KMVYEDTLTLFPFSGETVFMSENPGLMWILGCHNSDFRNGMTALLKVSSCDKNTGDYIE 720
DB 661 KMVYEDTLTLFPFSGETVFMSENPGLMWILGCHNSDFRNGMTALLKVSSCDKNTGDYIE 720
QY 721 DSYEDISAYLLSKNNAIBPRFSQNPVVKRQHREITRTTLQSDQEEIDYDDTISVMKK 780
DB 721 DSYEDISAYLLSKNNAIBPRFSQNPVVKRQHREITRTTLQSDQEEIDYDDTISVMKK 780
QY 781 EDFDIYDDENQSPRSFQKKTTHYFIAAVERLWDYGMSSPHVLNRQAQSVPOFKKVV 840
DB 781 EDFDIYDDENQSPRSFQKKTTHYFIAAVERLWDYGMSSPHVLNRQAQSVPOFKKVV 840
QY 841 FOEFTDGSFTOPLVRGELNEHLGLLGPYIRAEVDENIMVTFRNOASRPVSFVSSLSIYEE 900
DB 841 FOEFTDGSFTOPLVRGELNEHLGLLGPYIRAEVDENIMVTFRNOASRPVSFVSSLSIYEE 900
QY 901 DQROGAEPKKNFVKPNETKTYFWKVQHMAPTKDEFCCKAWAYFSDVLEKDVHSGLIGP 960
DB 901 DQROGAEPKKNFVKPNETKTYFWKVQHMAPTKDEFCCKAWAYFSDVLEKDVHSGLIGP 960
QY 961 LLVCHTNTLNPAHGROVTVQEFALFTTFIDETKSHYFTENMERNCRAPCNQIMEDPTPKE 1020
DB 961 LLVCHTNTLNPAHGROVTVQEFALFTTFIDETKSHYFTENMERNCRAPCNQIMEDPTPKE 1020
QY 1021 NYRFAINGYIMDTLPGLVMAQDQIRWYLLSMGSNENIHSIHPFSGHVTYVKKEEYKMA 1080
DB 1021 NYRFAINGYIMDTLPGLVMAQDQIRWYLLSMGSNENIHSIHPFSGHVTYVKKEEYKMA 1080
QY 1081 LYNLYPGVFETVEMLPKSGAGIWRVECLIGEHLHAGMSTFLVYSNKCOTPLGMAASHIRD 1140
DB 1081 LYNLYPGVFETVEMLPKSGAGIWRVECLIGEHLHAGMSTFLVYSNKCOTPLGMAASHIRD 1140
QY 1141 FQITASQYGOVAPKLABLHYSGSINAWSTKEPSPKVIKVDLLAPMIHGIKTQGAROKFS 1200
DB 1141 FQITASQYGOVAPKLABLHYSGSINAWSTKEPSPKVIKVDLLAPMIHGIKTQGAROKFS 1200
QY 1201 SLYTSQFIIMYSLDGKKWQTVRGSTGTLMVFFGNVDSGKHNIFNPPIARYIRLHPT 1260
DB 1201 SLYTSQFIIMYSLDGKKWQTVRGSTGTLMVFFGNVDSGKHNIFNPPIARYIRLHPT 1260
QY 1261 HYSIRSTLRMBELMGCDLNSCNPGLMESKASIDAQITASSYFTNNFATWSPSKARLHLQ 1320
DB 1261 HYSIRSTLRMBELMGCDLNSCNPGLMESKASIDAQITASSYFTNNFATWSPSKARLHLQ 1320
QY 1321 RSNARPOVNNPKEWLQVDFQKTMKVTVGTQGVKSLTSMYVKEFLISSQDGHQWTLF 1380
DB 1321 RSNARPOVNNPKEWLQVDFQKTMKVTVGTQGVKSLTSMYVKEFLISSQDGHQWTLF 1380
QY 1381 FQNGKVKVFOGNQDSFTPVVNSLOPPLTLRYLRHPQSWVHQIALRMEVLGCEAODLY 1438
DB 1381 FQNGKVKVFOGNQDSFTPVVNSLOPPLTLRYLRHPQSWVHQIALRMEVLGCEAODLY 1438
RESULT 2
AA21675
ID AAY21675 standard; Protein; 1457 AA.
XX AAY21675;
AC AAY21675;
DT 18-AUG-1999 (first entry)
XX Beta-domain deleted Factor VIII protein.
DE Factor VIII protein; gene modification; gene therapy; clinical disorder;
XX splicing pattern; RNA processing; gene regulation; beta-domain; human.
KW Homo sapiens.
OS
XX WO9929848-A1.
PN
XX 17-JUN-1999.
PD

XX PF 25-NOV-1998; 98NO-US25354.
XX PR 16-JAN-1998; 98US-0071596.
XX PR 05-DEC-1997; 97US-0067614.
XX PA (IMMU-) IMMUNE RESPONSE CORP.
XX PI Bidlingmaier S, Gonzales JEN, Ill CR, Yang CQ;
XX DR N-PSB; AAX82258, AAX82259, AAX82260.
XX DR WPI; 1999-385602/32.
XX PT Genes and vectors exhibiting increased expression and novel splicing
XX PT patterns, useful for expression of, e.g. beta-domain deleted factor
XX PT VIII
XX PS Disclosure; Page 72-78; 123pp; English.
XX CC The invention describes novel genes and vectors exhibiting increased
XX CC expression and novel splicing patterns. It provides a gene encoding a
XX CC Factor VIII protein, that comprises one or more consensus or near
XX CC consensus splice sites which have been corrected to increase expression.
XX CC The method, DNA sequences and expression vectors can be used to increase
XX CC the expression of a gene, especially a Factor VIII gene. Genes containing
XX CC modified 5' and/or 3' untranslated regions have optimized expression
XX CC levels and tissue-specific expression. The methods are used for
XX CC identification and correction of consensus splice sites, addition of
XX CC introns, optimization of 5' and 3' untranslated regions and increase in
XX CC cytoplasmic RNA accumulation. Hence the DNAs are useful in gene therapy
XX CC to treat a clinical disorder, to study RNA processing and/or gene
XX CC regulation. The present sequence represents a beta-domain deleted Factor
XX CC VIII protein.
XX SQ Sequence 1457 AA;

Query Match 100.0%; Score 7691; DB 20; Length 1457;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATRRYLGAVELSWDMQSDGLPVDARPPRPVPSFPFNTSVYKTLFVEFTVHLFN 60
DB 20 ATRRYLGAVELSWDMQSDGLPVDARPPRPVPSFPFNTSVYKTLFVEFTVHLFN 79
QY 61 IAKPRPWNGLLPTTQAEVYDVTWITLKNMASHPVSLHAGVSYWKASGEGAYDDQTSQ 120
DB 80 IAKPRPWNGLLPTTQAEVYDVTWITLKNMASHPVSLHAGVSYWKASGEGAYDDQTSQ 139
QY 121 REKEDDKVPFGSGHTYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
DB 140 REKEDDKVPFGSGHTYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199
QY 181 EGS LAKEKTQTLHKFTLLFAVDEGKSWSETKNSLMQDRDASARAWPKMHTVNGYVNR 240
DB 200 EGS LAKEKTQTLHKFTLLFAVDEGKSWSETKNSLMQDRDASARAWPKMHTVNGYVNR 259
QY 241 SLPGLIGCHRSYVHVIWGTTPEVHSIFLEGTFLVNRHQASLEISPIFLTAQTLL 300
DB 260 SLPGLIGCHRSYVHVIWGTTPEVHSIFLEGTFLVNRHQASLEISPIFLTAQTLL 319
QY 301 MDLGQFLPCHISSHQHDGMEAVKVDSCPEEPQLRMKNEEAEDYDDLTDSEMDVVRP 360
DB 320 MDLGQFLPCHISSHQHDGMEAVKVDSCPEEPQLRMKNEEAEDYDDLTDSEMDVVRP 379
QY 361 DDNSPSFTQIRSVAKHKTWVHYTAAREEDWDYAPLVLAPDDRYSKYQLNNGPQRIG 420
DB 380 DDNSPSFTQIRSVAKHKTWVHYTAAREEDWDYAPLVLAPDDRYSKYQLNNGPQRIG 439
QY 421 RYKVKRFWAYTDETFKTRAIQHESGILGPLLYGEGVDTLLIFKNQASRPNIYPHGI 480
DB 440 RYKVKRFWAYTDETFKTRAIQHESGILGPLLYGEGVDTLLIFKNQASRPNIYPHGI 499
QY 481 TDVRPLYSRRLPGKVHKLKDFPILPGEIFKYKWTVTVEGPTKSDPRCLTRYYSFVNME 540

DB 500 TDVRPLYSRRLPGKVHKLKDFPILPGEIFKYKWTVTVEGPTKSDPRCLTRYYSFVNME 559
QY 541 RDLASGLIGPLLIICYKESVDQGNQIMSKRNVILFSVFDENRSWYLTENIQRFLENPAG 600
DB 560 RDLASGLIGPLLIICYKESVDQGNQIMSKRNVILFSVFDENRSWYLTENIQRFLENPAG 619
QY 601 VOLEDEFOASNMHSINGYVFDLSQLSVCLHEVAYWYILSICGAQTDPLSVFSPGYTFKH 660
DB 620 VOLEDEFOASNMHSINGYVFDLSQLSVCLHEVAYWYILSICGAQTDPLSVFSPGYTFKH 679
QY 661 KMVYEDTLTLPFSGETFMENPCLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYIE 720
DB 680 KMVYEDTLTLPFSGETFMENPCLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYIE 739
QY 721 DSYEDISAYLLSKNNAIEPRSPQNPVLKRQREITRTTLOSDEIDYDDTISVEMKK 780
DB 740 DSYEDISAYLLSKNNAIEPRSPQNPVLKRQREITRTTLOSDEIDYDDTISVEMKK 799
QY 781 EDFDIYDEDNQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRNRAQSGVPQKVV 840
DB 800 EDFDIYDEDNQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRNRAQSGVPQKVV 859
QY 841 RQEFDTGSGTQPLYRGELNEHLGLLGPYIRAEVEDINMTFRNQASRPYSFSSLSIYBE 900
DB 860 RQEFDTGSGTQPLYRGELNEHLGLLGPYIRAEVEDINMTFRNQASRPYSFSSLSIYBE 919
QY 901 DORQAGPEPKNFVKNETKTYFWKHMAPTKDFDKAWAYFSDVLEKDVHSLGIGP 960
DB 920 DORQAGPEPKNFVKNETKTYFWKHMAPTKDFDKAWAYFSDVLEKDVHSLGIGP 979
QY 961 LLVCHTNTLPAHGQVTVQEPALFPTTIDETKSMYFTENMERNCRAPCNIQMEDPTFKE 1020
DB 980 LLVCHTNTLPAHGQVTVQEPALFPTTIDETKSMYFTENMERNCRAPCNIQMEDPTFKE 1039
QY 1021 NYRFAINGYIMDTLPLVMAQDQIRWYLLSGNSNENIHSIFSGHVTVRKKEBKMA 1080
DB 1040 NYRFAINGYIMDTLPLVMAQDQIRWYLLSGNSNENIHSIFSGHVTVRKKEBKMA 1099
QY 1081 LYNLPVGVETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGWASGHID 1140
DB 1100 LYNLPVGVETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGWASGHID 1159
QY 1141 FOITASGOYGOWAPKLARLHYSGSINANSTKEPFSWIKVDLAPMLIHGIKTOGARQKFS 1200
DB 1160 FOITASGOYGOWAPKLARLHYSGSINANSTKEPFSWIKVDLAPMLIHGIKTOGARQKFS 1219
QY 1201 SLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSGIGKHINFPNPIIARYIRLHPT 1260
DB 1220 SLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSGIGKHINFPNPIIARYIRLHPT 1279
QY 1261 HYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMPATWSPSKARLHLOG 1320
DB 1280 HYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMPATWSPSKARLHLOG 1339
QY 1321 RSNAMRPQVNNPKEMQLQVDFQTKMVTGVTTOGVKSLLTSMYKBEFLISSQDGHQWTLF 1380
DB 1340 RSNAMRPQVNNPKEMQLQVDFQTKMVTGVTTOGVKSLLTSMYKBEFLISSQDGHQWTLF 1399
QY 1381 FQNGKVPQGNQDSFTPVVNSLDPLLTRYLRIHPQSWVHQAIALRMEVLGCEAQDLY 1438
DB 1400 FQNGKVPQGNQDSFTPVVNSLDPLLTRYLRIHPQSWVHQAIALRMEVLGCEAQDLY 1457

RESULT 3
AAW46246
ID AAW46246 standard; Protein; 1457 AA.
XX
AC AAW46246;
XX
DT 25-MAR-2003 (updated)
DT 06-AUG-1998 (first entry)
XX

Human factor VIII beta-domain deleted SQN deletion protein sequence.

DE XX Replication defective; recombinant retrovirus; RRV; therapeutic protein;
KW haemophilia; thrombosis; hypercoagulable disorder; liver disease; human;
KW hepatitis; thalassemia; phenylketonuria; Lesch-Nyhan syndrome; diabetes;
KW cystic fibrosis; Duchenne's Muscular Dystrophy; hypercholesterolemia;
KW hypopituitarism; adenine deaminase deficiency; HIV infection; anaemia;
KW Guacher's syndrome; high blood pressure; Alzheimer's disease; autoimmune;
KW inflammatory disease; factor VIII.

OS Homo sapiens.

PN WO9800541-A2.

XX 08-JAN-1998.

XX 02-JUL-1997; 97WO-US11784.

XX 04-JUN-1997; 97US-0869309.

PR 03-JUN-1996; 96US-0645601.

PR 13-AUG-1996; 96US-0696381.

PA (CHIR) CHIRON CORP.

PI Jolly DJ, Barber JR, Chang SMW, Respass JG, Allen JR, Boder M;

PI Chong K, De La Vega D, Depolo NJ, Hsu DC, Ibanez CE;

PI Mittelstaedt DM, Pruseak CE, Greengard J, Lee R;

XX WPI; 1998-086966/08.

XX N-PSDB; AAV19581.

PT New replication defective recombinant retro-viruses - which can be
PT administered to provide long term systemic expression of therapeutic
PT protein in blood, useful in, e.g. treating hyper-coagulable
PT disorders

XX Example 28; Pages 213-217; 272pp; English.

CC This is the beta-domain deleted SQN deletion protein of human factor
CC VIII. The encoding DNA is used to construct recombinant retroviral
CC vectors expressing human factor VIII. The invention provides the
CC preparation of replication defective recombinant retrovirus (RRV)
CC expressing a therapeutic protein. The RRV preparation is resistant to
CC degradation by human complement and is capable of inducing long term
CC systemic expression of the therapeutic protein when administered
CC intravenously to a human. The long term systemic expression results in a
CC measurable level of the therapeutic protein being produced in the blood
CC of the human for a period of at least 30 days after the administration of
CC the RRV vector preparation. RRV's can be used for in vivo delivery of
CC therapeutic protein to treat, e.g. haemophilia A, haemophilia B,
CC thrombosis, hypercoagulable disorders, liver diseases such as hepatitis,
CC disorders such as thalassemia, phenylketonuria, Lesch-Nyhan syndrome,
CC severe combined immunodeficiency (SCID), cystic fibrosis, Duchenne's
CC Muscular Dystrophy, inherited emphysema, familial hypercholesterolemia,
CC diabetes, hypopituitarism, adenine deaminase deficiency, alphas-
CC antitrypsin deficiency, Guacher's syndrome, anaemia, infections such as
CC HIV infection, high blood pressure, Alzheimer's disease, autoimmune or
CC inflammatory disease or graft versus host disease. RRV's are capable of
CC surviving inactivation in human serum thereby allowing efficient gene
CC transfer over prolonged periods of time.
CC (Updated on 25-MAR-2003 to correct PI field.)

SQ Sequence 1457 AA;

Query Match 99.9%; Score 7684; DB 19; Length 1457;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1437; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATRRYLGAVELSWDYMOSDLGELPVDARPPRPVKSPFNTSVVYKKTFLVFETVHLFN 60

Db 20 ATRRYLGAVELSWDYMOSDLGELPVDARPPRPVKSPFNTSVVYKKTFLVFETVHLFN 79

Qy 61 IAKPRPFWMLLGGFTIOAEVYDVTVVITLKNMASHPVSLHVGVSYWKASGAEYDDQTSQ 120

Db 80 IAKPRPFWMLLGGFTIOAEVYDVTVVITLKNMASHPVSLHVGVSYWKASGAEYDDQTSQ 139
Qy 121 REKEDDKVFGGSHYTVQVLKENGPMASDPLCLTYSYLSHVLDVLDNSLGLGALLVCR 180
Db 140 REKEDDKVFGGSHYTVQVLKENGPMASDPLCLTYSYLSHVLDVLDNSLGLGALLVCR 199
Qy 181 EGSIAKEKTQTLHKFILLFAVFDGEGSWSETKNSLMQDRDAASARAAPPGHVTNGVNR 240
Db 200 EGSIAKEKTQTLHKFILLFAVFDGEGSWSETKNSLMQDRDAASARAAPPGHVTNGVNR 259
Qy 241 SLPGLIGCHRSVYVHVGITGTPVHSLFLEGTFLVRNHRQASLEISITFLTAQTLL 300
Db 260 SLPGLIGCHRSVYVHVGITGTPVHSLFLEGTFLVRNHRQASLEISITFLTAQTLL 319
Qy 301 MDLQGQFLLFCHISSHQHDGMEAYVKVDSCEPEEPQLRMKNNEEAEDVDDDLTDSMDVVR 360
Db 320 MDLQGQFLLFCHISSHQHDGMEAYVKVDSCEPEEPQLRMKNNEEAEDVDDDLTDSMDVVR 379
Qy 361 DDNSPFSFIQIRSVAKKHGPKTWVHYIAAEEEDWDYAPLVLPADDRSYKSYLNNGPORIG 420
Db 380 DDNSPFSFIQIRSVAKKHGPKTWVHYIAAEEEDWDYAPLVLPADDRSYKSYLNNGPORIG 439
Qy 421 RYKVKVREWAYTDETFTKTRAIQHESGILGPLLYGEVGDITLLIIFKNQASRPYNIYPHGI 480
Db 440 RYKVKVREWAYTDETFTKTRAIQHESGILGPLLYGEVGDITLLIIFKNQASRPYNIYPHGI 499
Qy 481 TDVRPLYSRRLPGVXKHLKDFPILPGEIFKYKKTWTVVDEGPTKSDPRCLTRYIYSSFNME 540
Db 500 TDVRPLYSRRLPGVXKHLKDFPILPGEIFKYKKTWTVVDEGPTKSDPRCLTRYIYSSFNME 559
Qy 541 RDLASGLIGPILLICYESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQFLPNPAG 600
Db 560 RDLASGLIGPILLICYESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQFLPNPAG 619
Qy 601 VQLEDEPFOASNIWHSINGVYVDSLOLSVCLHVAWYIILSIGAQTDFLSVFFSGYTFKH 660
Db 620 VQLEDEPFOASNIWHSINGVYVDSLOLSVCLHVAWYIILSIGAQTDFLSVFFSGYTFKH 679
Qy 661 KMVYEDTLTLFPFSGETVFMSENPGWLILGCHNSDFRNRGTMALLKVSSCDXNTGYDE 720
Db 680 KMVYEDTLTLFPFSGETVFMSENPGWLILGCHNSDFRNRGTMALLKVSSCDXNTGYDE 739
Qy 721 DSVEDISAYLLSKNNAIEPRFSQNPVPLKRGHREITRTTLQSDQBEIDYDDTISVEMKK 780
Db 740 DSVEDISAYLLSKNNAIEPRFSQNPVPLKRGHREITRTTLQSDQBEIDYDDTISVEMKK 799
Qy 781 EDFDIYDEDENQSFRSQKTRHYFTAAVERLDYGMSSSPHYLRNRAQSGSVPOPKKV 840
Db 800 EDFDIYDEDENQSFRSQKTRHYFTAAVERLDYGMSSSPHYLRNRAQSGSVPOPKKV 859
Qy 841 FQEFDTGSGFTQPLRYGELNEHLGLLGPYIRAEVEDNIMVTFRNOASRPYSFYSLSLYEE 900
Db 860 FQEFDTGSGFTQPLRYGELNEHLGLLGPYIRAEVEDNIMVTFRNOASRPYSFYSLSLYEE 919
Qy 901 DQRCQAEPRKNFKVKNETKTYFMVQHMAPTKDEFCCKAWAYFSDVLEKOVHSGLIGP 960
Db 920 DQRCQAEPRKNFKVKNETKTYFMVQHMAPTKDEFCCKAWAYFSDVLEKOVHSGLIGP 979
Qy 961 LLVCHNTNLNPAHGRQVTVQEFALPFTIDEKTSWYTFENMERNCRAPCNIQMEDPTFKE 1020
Db 980 LLVCHNTNLNPAHGRQVTVQEFALPFTIDEKTSWYTFENMERNCRAPCNIQMEDPTFKE 1039
Qy 1021 NYRFAHNGYIMDTLPLGLVMAQDQRIRWYLLSGNSNENIHSIFHSGHVFVTRKKEEYKMA 1080
Db 1040 NYRFAHNGYIMDTLPLGLVMAQDQRIRWYLLSGNSNENIHSIFHSGHVFVTRKKEEYKMA 1099
Qy 1081 LYNLYPGVFETVEMLPFSKAGINREVECLIGEHLAGHAGTFLVYSNKCQTPGLGNASGHIRD 1140
Db 1100 LYNLYPGVFETVEMLPFSKAGINREVECLIGEHLAGHAGTFLVYSNKCQTPGLGNASGHIRD 1159
Qy 1141 FQITAGSQYQWAPKILARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQCARQKFS 1200

Db 1160 FQITASGOYQWAPKLARLHYSGINAWSTKEPFSWIKVDLLAPMIHGIKTOGAROKFS 1219
 Qy 1201 SLYISQFIIMYSLDGKKQTYRGNSGTGLMVFVFGNVDSSGKHNIFNPPIIARIHLPT 1260
 Db 1220 SLYISQFIIMYSLDGKKQTYRGNSGTGLMVFVFGNVDSSGKHNIFNPPIIARIHLPT 1279
 Qy 1261 HYSIRSTRIRWELMCGDLNCSMPGMSKASDAQITASSYFTNMFATWSPSKARLHLQ 1320
 Db 1280 HYSIRSTRIRWELMCGDLNCSMPGMSKASDAQITASSYFTNMFATWSPSKARLHLQ 1339
 Qy 1321 RSNARWPOVNNPKWLOVDFOKTMKVTGVTGQVKSLTSMYVKEFLISSQDGHQWTLF 1380
 Db 1340 RSNARWPOVNNPKWLOVDFOKTMKVTGVTGQVKSLTSMYVKEFLISSQDGHQWTLF 1399
 Qy 1381 FONGKVKVQGNQDSFTPVNSLDPPLTRYIRIHQSWHQAIALRMEVLGCEAQDLY 1438
 Db 1400 FONGKVKVQGNQDSFTPVNSLDPPLTRYIRIHQSWHQAIALRMEVLGCEAQDLY 1457

RESULT 4

AAW44372
 ID AAW44372 standard; Protein; 1457 AA.

AC AAW44372;

XX 20-JUL-1998 (first entry)

XX Human Factor VIII SQN deletion mutant.

XX Factor VIII; blood clotting; haemophilia A; gene therapy;

XX retrovirus; vector; human.

XX Homo sapiens.

XX Synthetic.

XX W09800542-A2.

XX 08-JAN-1998.

XX 02-JUL-1997; 97WO-US11785.

XX 04-JUN-1997; 97US-0869309.

XX 03-JUL-1996; 96US-0645601.

XX 13-AUG-1996; 96US-0696381.

XX (CHIR) CHIRON CORP.

XX Allen JR, Barber JR, Boder M, Chang SM, Chong K;

XX De La Vega D, Depolo NJ, Greengard J, Hsu DC, Ibanez CE;

XX Jolly DJ, Mittelstaedt DM, Prussak CE, Respass JG;

XX WPI; 1998-086967/08.

XX N-PSDB; AAV15338.

XX New replication defective recombinant retroviruses - which express B

XX domain-deleted human factor VIII or human factor IX for the

XX treatment of haemophilia

XX Claim 5; Page 175-180; 236pp; English.

CC This polypeptide comprises the B domain deletion mutant SQN of
 CC human Factor VIII. The SQN mutant is created by fusing Ser-743 to
 CC Gln-1638 of native Factor VIII (see AAW44373) to form a Ser-Gln-Aan
 CC (SQN) link between the A2 and A3 Factor VIII domains. A DNA
 CC sequence encoding the SQN deletion mutant is provided in AAV15338.
 CC When compared to plasmid-derived Factor VIII, the SQN deletion does
 CC not influence the in vivo pharmacokinetics, but the reduced size of
 CC the molecule appears to decrease proteolytic degradation. The
 CC invention relates to preparations of replication defective
 CC recombinant retrovirus (RV) expressing a B domain-deleted human
 CC Factor VIII protein, where the recombinant RV is capable of
 CC infecting human cells, is resistant to degradation by human
 CC complement and is capable of inducing long-term (at least 30 days

CC and up to 6 months or longer post-injection) systemic expression of
 CC Factor VIII when administered to a haemophilia A patient.

SQ Sequence 1457 AA;

Query Match 99.9%; Score 7684; DB 19; Length 1457;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1437; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATRRYIGAVELSDVMQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 60

Db 20 ATRRYIGAVELSDVMQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 79

Qy 61 IAKPRPPWMLGLGPTTIAQEVYDVITLKNMASHPVSLHAGVSYWKASEGAEYDDDTSQ 120

Db 80 IAKPRPPWMLGLGPTTIAQEVYDVITLKNMASHPVSLHAGVSYWKASEGAEYDDDTSQ 139

Qy 121 REKEDDKVFFGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180

Db 140 REKEDDKVFFGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199

Qy 181 EGSIAKEKTTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVGVNVR 240

Db 200 EGSIAKEKTTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVGVNVR 259

Qy 241 SLPLGLIGHCRKSVYVHVGITTPVHSHIFLEGHTPLVRNHRQASLEISPIITLTAQTLL 300

Db 260 SLPLGLIGHCRKSVYVHVGITTPVHSHIFLEGHTPLVRNHRQASLEISPIITLTAQTLL 319

Qy 301 MDLGQFLFCHISSHQHDGMEAYVKVDSCEEPQLRMKNNEAEYDDDLTDSMDVVR 360

Db 320 MDLGQFLFCHISSHQHDGMEAYVKVDSCEEPQLRMKNNEAEYDDDLTDSMDVVR 379

Qy 361 DDNSPSFIQIRSVAKKPKTWHTYIAAEEDWDYAPLVLPDDRYSKSYQLNNGPORIG 420

Db 380 DDNSPSFIQIRSVAKKPKTWHTYIAAEEDWDYAPLVLPDDRYSKSYQLNNGPORIG 439

Qy 421 RYKVKRFMAYTDETFKTREAIQHEGILGLPLYGEVGDTLIIIFKNQASRPYNIYPHGI 480

Db 440 RYKVKRFMAYTDETFKTREAIQHEGILGLPLYGEVGDTLIIIFKNQASRPYNIYPHGI 499

Qy 481 TDVRPLYSRRLPGVKVHLKDFPILPGEIFKYIKWTVTVEDEGPTKSDPRCLTRYISSFVNME 540

Db 500 TDVRPLYSRRLPGVKVHLKDFPILPGEIFKYIKWTVTVEDEGPTKSDPRCLTRYISSFVNME 559

Qy 541 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRPFPNAG 600

Db 560 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRPFPNAG 619

Qy 601 VQLEDPEFOASNMHSINGVYVDSLQLSVCLHEVAYWYILSIGAQTDFLSVFPSSGYTFKH 660

Db 620 VQLEDPEFOASNMHSINGVYVDSLQLSVCLHEVAYWYILSIGAQTDFLSVFPSSGYTFKH 679

Qy 661 KMVYEDTLTLFPFSGETVFMNSMENPGLIILGCHNSDFRNRMGTALLKVSSCDXNTGDIYE 720

Db 680 KMVYEDTLTLFPFSGETVFMNSMENPGLIILGCHNSDFRNRMGTALLKVSSCDXNTGDIYE 739

Qy 721 DSYEDISAYLLSKNNAIEPRSPQNPPVLKRQREITRTTLQSDQBEIDYDDTISVEMKK 780

Db 740 DSYEDISAYLLSKNNAIEPRSPQNPPVLKRQREITRTTLQSDQBEIDYDDTISVEMKK 799

Qy 781 EDFDIYDEBENQSPRSFQKTRHYFTAAVERLDWYGMSSSPHYLRNRAQSGSVQPKVV 840

Db 800 EDFDIYDEBENQSPRSFQKTRHYFTAAVERLDWYGMSSSPHYLRNRAQSGSVQPKVV 859

Qy 841 FQEFDTGSGFTQPLYRGELNEHGLGIPYTRAEVEDNMVTFRNQASRPYSFYSSLSIYEE 900

Db 860 FQEFDTGSGFTQPLYRGELNEHGLGIPYTRAEVEDNMVTFRNQASRPYSFYSSLSIYEE 919

Qy 901 DQRQGAEPKRNFKVKNETKTYFMKVOHMAPTKDEFDCKAWYFSDVDLEKQVHSGLIGP 960

Db 920 DQRQGAEPKRNFKVKNETKTYFMKVOHMAPTKDEFDCKAWYFSDVDLEKQVHSGLIGP 979

QY 961 LLVCHTNTLNPAHGRQVTVQEPALFTTIFDETKSWYFTENMERNCRAPCNIOWEDPTFK 1020
Db 980 LLVCHTNTLNPAHGRQVTVQEPALFTTIFDETKSWYFTENMERNCRAPCNIOWEDPTFK 1039
QY 1021 NYRFAHNGIYIMDTLPGLVMAQDQRIWYLLSGNSNENIHSIFSGHVFTVRKCEYKMA 1080
Db 1040 NYRFAHNGIYIMDTLPGLVMAQDQRIWYLLSGNSNENIHSIFSGHVFTVRKCEYKMA 1099
QY 1081 LYNLYPGVETVEMLPKAGINRVECLIGEHLHAGNSTLFLVYSNKCQTPPLGWSGHIRD 1140
Db 1100 LYNLYPGVETVEMLPKAGINRVECLIGEHLHAGNSTLFLVYSNKCQTPPLGWSGHIRD 1159
QY 1141 FOITASQGOQWAPKLARLHYSGINAMSTKPFWSIKVDLAPMIHGIKQGAROKFS 1200
Db 1160 FOITASQGOQWAPKLARLHYSGINAMSTKPFWSIKVDLAPMIHGIKQGAROKFS 1219
QY 1201 SLVYSQFIIMYSIDGKKWQTYRGNSTGTLWFFGNVDSGIKHNIFNPPIIARYLHPT 1260
Db 1220 SLVYSQFIIMYSIDGKKWQTYRGNSTGTLWFFGNVDSGIKHNIFNPPIIARYLHPT 1279
QY 1261 HYSIRSTLRMELMGCDLNSCSMPLGWSKAISSAQITASSYFTNMPATWSPSKARLHQQ 1320
Db 1280 HYSIRSTLRMELMGCDLNSCSMPLGWSKAISSAQITASSYFTNMPATWSPSKARLHQQ 1339
QY 1321 RSNARPOVNNPKEWLOVDFOKTMKVTGTTQGVKSLLTSMYKSEFLSSSQDGHQWTLF 1380
Db 1340 RSNARPOVNNPKEWLOVDFOKTMKVTGTTQGVKSLLTSMYKSEFLSSSQDGHQWTLF 1399
QY 1381 FQNGKVKVQGNQDSFTPVVNSLDPLLTRYLRIHPQSWHQAIALRMEVLGCEAODLY 1438
Db 1400 FQNGKVKVQGNQDSFTPVVNSLDPLLTRYLRIHPQSWHQAIALRMEVLGCEAODLY 1457

RESULT 5
AAW23414
ID AAW23414 standard; Protein; 1471 AA.
XX
AC AAW23414;
XX

XX 08-APR-1998 (first entry)
XX Human B-domain deleted factor VIII protein.

XX Post-translational regulatory element; PRE; enhancer II; intronless gene;
XX surface antigen gene; cytoplasmic accumulation; targeted delivery;
XX near consensus splice sequence; blood coagulation factor; factor VIII;
XX factor IX.

XX Homo sapiens.
XX
XX WO9733994-A1.
XX
XX 18-SEP-1997.
XX
XX 10-MAR-1997; 97WO-US03561.
XX
XX 11-MAR-1996; 96US-0683839.
XX
XX (IMMU-) IMMUNE RESPONSE CORP.
XX
XX Bidlingmaier S, Ill CR;
XX
XX WPI; 1997-470874/43.
XX
XX N-PSDB; AAT73164.
XX

XX Vector for increased expression of intronless genes - comprises
XX intronless gene with at least one near consensus splice sequence, a
XX promoter and at least one viral cis-acting post-transcriptional
XX regulatory element

XX Example 1; Pages 31-36; 59pp; English.

XX The present sequence represents human B-domain deleted factor VIII

CC The cDNA encoding this protein also contains, 3' of the coding region, a
CC post-translational regulatory element (PRE) of the Hepatitis B virus.
CC PRE sequences have been shown to function in cis to increase the
CC steady-state levels of surface gene transcripts by facilitating
CC cytoplasmic accumulation of these transcripts. The above nucleic acid
CC sequence is part of a novel vector, comprising an intronless gene
CC containing 1 or more near consensus splice sequences operably linked to
CC a promoter sequence so that the gene is transcribed in a cell.
CC Intronless gene transcripts which contain near consensus splice site
CC sequences are believed to get tied up in the nucleus of the cell where
CC splicing occurs, rather than being transported to the cytoplasm where
CC they can be translated into proteins. The PRE sequences are transcribed
CC along with the gene, causing export of the gene transcript from the
CC nucleus into the cytoplasm of the cell. The vector can be used
CC to increase the expression of an intronless gene containing at least one
CC near consensus splice sites, preferably cDNA encoding a blood coagulation
CC factor, particularly Factor VIII or IX. The complex allows the targeted
CC delivery of the vector to a specific cell, e.g. hepatocytes when the
CC ligand is an asialoglycoprotein which binds the asialoglycoprotein
CC receptor present on their surface.

XX Sequence 1471 AA;

Query Match 99.8%; Score 7674; DB 18; Length 1471;

Best Local Similarity 99.0%; Pred. No. 0;

Matches 1438; Conservative 0; Mismatches 0; Indels 14; Gaps 1;

QY 1 ATRRYLGAVELSDWDMQSDLGELPVDARFPFRVPKSPFFNTSVVYKKTFLVEFTVHLFN 60
Db 20 ATRRYLGAVELSDWDMQSDLGELPVDARFPFRVPKSPFFNTSVVYKKTFLVEFTVHLFN 79
QY 61 IAKPRPPMGLLGPTIQAEVYDVTVTILKNWASHPSVLHVGVSYSWKAEGAYDDQTSQ 120
Db 80 IAKPRPPMGLLGPTIQAEVYDVTVTILKNWASHPSVLHVGVSYSWKAEGAYDDQTSQ 139
QY 121 REKEDDKVFGSGSHYVWVLKENGPMASDPLCLTYSYLSHVLDLKDLSGLIGALLVCR 180
Db 140 REKEDDKVFGSGSHYVWVLKENGPMASDPLCLTYSYLSHVLDLKDLSGLIGALLVCR 199
QY 181 EGS�AKEKTQTLHKFILLFAVFDGKSWHSETKSNLMQDRDAASARAPKQHTVNGVYNR 240
Db 200 EGS�AKEKTQTLHKFILLFAVFDGKSWHSETKSNLMQDRDAASARAPKQHTVNGVYNR 259
QY 241 SLPLGLICHRKSVYWHVIGMTTPFVHSIFLEGTFLVRNHRQASLEISPIITFLTAOTLL 300
Db 260 SLPLGLICHRKSVYWHVIGMTTPFVHSIFLEGTFLVRNHRQASLEISPIITFLTAOTLL 319
QY 301 MDLQGFLLFCHISSHQHDGMEAYVKVDSCPBPOLRMKNNEAEYDDDLTDSMDVVRP 360
Db 320 MDLQGFLLFCHISSHQHDGMEAYVKVDSCPBPOLRMKNNEAEYDDDLTDSMDVVRP 379
QY 361 DDNSPSFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLVLAAPDDRYSKQSVLNNGPQIRG 420
Db 380 DDNSPSFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLVLAAPDDRYSKQSVLNNGPQIRG 439
QY 421 RYKVKRPMAYTDEFTKTREAIQHESGILGPLLYGVEGDTLLIIFKNQASRPYNIYPHGI 480
Db 440 RYKVKRPMAYTDEFTKTREAIQHESGILGPLLYGVEGDTLLIIFKNQASRPYNIYPHGI 499
QY 481 TDVRPLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGPKDPRCLITRYSFVWME 540
Db 500 TDVRPLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGPKDPRCLITRYSFVWME 559
QY 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDNRNRSWYLTENIQRLPNPAG 600
Db 560 RDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDNRNRSWYLTENIQRLPNPAG 619
QY 601 VQLEDPEFQASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDPLSFVFFSGYTPKH 660
Db 620 VQLEDPEFQASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDPLSFVFFSGYTPKH 679
QY 661 KMWVEDTLTLPFPFSGETVFMSENPGWLWILGCHNSDFNRGWTALLKVSSCDKNGDYIE 720

Db 680 KMVETDLTLFPFGSETVFMSENPGLWILGCHNSDFNRGMGALLKVVSCDKNKTGDYYE 739
QY 721 DSYEDISAYLLSKNAIEPRFSQN-----PPVLKRHQREITRTTLOSQDE 766
Db 740 DSYEDISAYLLSKNAIEPRFSQNSRHPSTROKQFNATPPVLKRHQREITRTTLOSQDE 799
QY 767 EIDYDDTISVEMKEDPDIIDENQSPRQFKKTRHYFIAAVERLWDYOMSSSPHVLRN 826
Db 800 EIDYDDTISVEMKEDPDIIDENQSPRQFKKTRHYFIAAVERLWDYOMSSSPHVLRN 859
QY 827 RAQSGSVPOPKVVFORFTDGSFTQPLRYRGELNEHLGLGPIYRAAEVDNIMWTERNOAS 886
Db 860 RAQSGSVPOPKVVFORFTDGSFTQPLRYRGELNEHLGLGPIYRAAEVDNIMWTERNOAS 919
QY 887 RPYSFYSSLSIYSEEDQKGAEPKRNFKVKNETKTYFWKHMAPTKDEDFCKAWAYFSD 946
Db 920 RPYSFYSSLSIYSEEDQKGAEPKRNFKVKNETKTYFWKHMAPTKDEDFCKAWAYFSD 979
QY 947 VDLEKDVHSLIGLPLLVCHTNTLNPAGROVTVQEFALFTTIDETKSWYFTENMERNCR 1006
Db 980 VDLEKDVHSLIGLPLLVCHTNTLNPAGROVTVQEFALFTTIDETKSWYFTENMERNCR 1039
QY 1007 APCNIQMEDPTFKENRFHAIYIMDTLPLGVMAQDORIRWYLLSMGNSNENIHSIFSG 1066
Db 1040 APCNIQMEDPTFKENRFHAIYIMDTLPLGVMAQDORIRWYLLSMGNSNENIHSIFSG 1099
QY 1067 HVFTVRKKEEYKMAVNLXPGVZETVEMLPSPKAGINRVECLIGEHLHAGMSTLFLVYSNK 1126
Db 1100 HVFTVRKKEEYKMAVNLXPGVZETVEMLPSPKAGINRVECLIGEHLHAGMSTLFLVYSNK 1159
QY 1127 CQPLGMASGHIRDFOITASQYQGWAPKARLHYSGSINAWSTKEPFSWIKVDLLAPMI 1186
Db 1160 CQPLGMASGHIRDFOITASQYQGWAPKARLHYSGSINAWSTKEPFSWIKVDLLAPMI 1219
QY 1187 IHGKTKGAKQKSSLYISOFIIMYSIDGKKWQTYRGNSGTLMVRFGNVDSSGIRHNIF 1246
Db 1220 IHGKTKGAKQKSSLYISOFIIMYSIDGKKWQTYRGNSGTLMVRFGNVDSSGIRHNIF 1279
QY 1247 NPPIIARYIRLHPHTHYSIRTLRMLMGCDLNSCMPLGMSKAISSAQITASSYFTNMF 1306
Db 1280 NPPIIARYIRLHPHTHYSIRTLRMLMGCDLNSCMPLGMSKAISSAQITASSYFTNMF 1339
QY 1307 ATWSPSKARLHLQGRSNARPPQVNNPKEWLQVDFQKTMKVTVGTQGVKSLLTSMYKKEF 1366
Db 1340 ATWSPSKARLHLQGRSNARPPQVNNPKEWLQVDFQKTMKVTVGTQGVKSLLTSMYKKEF 1399
QY 1367 LISSQDGHQWTLFFQNGKVKVFGNQDSTFPVNSLDPLLRILYLRHQPQSWHQTALR 1426
Db 1400 LISSQDGHQWTLFFQNGKVKVFGNQDSTFPVNSLDPLLRILYLRHQPQSWHQTALR 1459
QY 1427 MEVLGCEAQDLY 1438
Db 1460 MEVLGCEAQDLY 1471

RESULT 6

AAB67959
ID AAB67959 standard; Protein; 1471 AA.

XX
XX
AC AAB67959;

XX
XX

DT 29-JUN-2001 (first entry)
XX
XX

DE Amino acid sequence of human B-domain deleted factor VIII.
XX

XX Adeno-associated virus vector; B-domain; factor VIII; haemophilia A;
XX coagulation disorder.
XX

OS Homo sapiens.
XX

XX
XX W0200127303-A1.

XX
XX 19-APR-2001.

XX 12-OCT-2000; 2000WO-US28221.
XX
XX 12-OCT-1999; 99US-0158780.
XX
XX (UYNC-) UNIV NORTH CAROLINA.
XX
XX Walsh CE, Chao H, Burstein H, Lynch CM, Stepan AM, Munson K;
XX
XX MPI; 2001-273781/28.
XX
XX N-PSDB; AAR64647.
XX
XX New recombinant adeno-associated virus vector, useful for treating
XX haemophilia A, comprises heterologous nucleotide sequence encoding
XX B-domain deleted human factor VIII operably linked with liver-preferred
XX expression control element -
XX
XX Example 1; Fig 1; 87pp; English.
XX
XX The specification describes a recombinant adeno-associated virus (rAAV)
XX vector. The vector comprises a heterologous nucleotide sequence
XX encoding B-domain deleted factor VIII operably linked with at least one
XX enhancer and at least one promoter. The method results in the production
XX of high titer rAAV vector stocks carrying the B-domain deleted factor
XX VIII transgenes and expression cassettes, which generate adequate titers
XX of virus for in vivo administration. The recombinant vectors are useful
XX for treating haemophilia A, where the liver expresses the encoded
XX B-domain deleted factor VIII, which is secreted into the blood. They are
XX also useful for the treatment of other coagulation disorders. The
XX present sequence represents a B-domain deleted factor VIII.
XX
XX Sequence 1471 AA;
XX
XX Query Match 99.8%; Score 7674; DB 22; Length 1471;
XX Best Local Similarity 99.0%; Pred.No. 0;
XX Matches 1438; Conservative 0; Mismatches 0; Indels 14; Gaps 1;
XX
QY 1 ATRRYLGVAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVYKTLFVEFTVHLFN 60
Db 20 ATRRYLGVAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVYKTLFVEFTVHLFN 79
QY 61 IAKPRPFWMLGPTTQAEVYDVTWITLKNMASHPVSLHAGVSYNKASGAYDDQTSQ 120
Db 80 IAKPRPFWMLGPTTQAEVYDVTWITLKNMASHPVSLHAGVSYNKASGAYDDQTSQ 139
QY 121 REKEDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLTALLVCR 180
Db 140 REKEDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLTALLVCR 199
QY 181 EGSIAKEKTQTLHKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVNR 240
Db 200 EGSIAKEKTQTLHKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVNR 259
QY 241 SLPGLIGCHRKSVYMHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTITLTAQTL 300
Db 260 SLPGLIGCHRKSVYMHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTITLTAQTL 319
QY 301 MDLGQFLFCHISSHQHDGMEAVKVDSCPEEPQLRMKNNEAEYDDDLTDSMDVVR 360
Db 320 MDLGQFLFCHISSHQHDGMEAVKVDSCPEEPQLRMKNNEAEYDDDLTDSMDVVR 379
QY 361 DDNSPSFIQIRSVAKKHPTWVHYIAABEEDWDYAPLVLPDDRYSKSOYLANGPQIRIG 420
Db 380 DDNSPSFIQIRSVAKKHPTWVHYIAABEEDWDYAPLVLPDDRYSKSOYLANGPQIRIG 439
QY 421 RYKVKRFWAYTDETPKTRAIQHSIGLGPLYGVGDTLLIFKNQASRPNIYPHGI 480
Db 440 RYKVKRFWAYTDETPKTRAIQHSIGLGPLYGVGDTLLIFKNQASRPNIYPHGI 499
QY 481 TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEGPTKSDPRCLTRYSSFFVME 540
Db 500 TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEGPTKSDPRCLTRYSSFFVME 559

Qy 541 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSHWLTENTQRELPNDAG 600
Db 560 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSHWLTENTQRELPNDAG 619
Qy 601 VOLEDPEFQASNIHMSINGYVFDLSQLSVCLHEVAYWILSIGACTDFLSFFSGYTPKH 660
Db 620 VOLEDPEFQASNIHMSINGYVFDLSQLSVCLHEVAYWILSIGACTDFLSFFSGYTPKH 679
Qy 661 KMVYEDTLTLFPFGSETVFMENPGLWILGCHNSDFRNQMTALLKVSSCDKNTGDIYE 720
Db 680 KMVYEDTLTLFPFGSETVFMENPGLWILGCHNSDFRNQMTALLKVSSCDKNTGDIYE 739
Qy 721 DSYEDISAYLSKNAIIEPRFSQY-----PPVLKHQREIRTRTLQSDQE 766
Db 740 DSYEDISAYLSKNAIIEPRFSQNSRHPSTKQFNATPPVLKHQREIRTRTLQSDQE 799
Qy 767 EIDYDDTTLVEMKEDFDIYDENQSPRSFQKTRHYFIAAVERLWDYGMSSPHVLRN 826
Db 800 EIDYDDTTLVEMKEDFDIYDENQSPRSFQKTRHYFIAAVERLWDYGMSSPHVLRN 859
Qy 827 RAQSGSVQFKKVVQEFQFTDGSFTQPLRYGELNEHLGLLGPYIRAEVEDNIMVTFRNQAS 886
Db 860 RAQSGSVQFKKVVQEFQFTDGSFTQPLRYGELNEHLGLLGPYIRAEVEDNIMVTFRNQAS 919
Qy 887 RPYSPYSSLSIYEEHQROGAPRKNFVQPNETKTYFWKVQHHMPTKDEPCKAWAYFSD 946
Db 920 RPYSPYSSLSIYEEHQROGAPRKNFVQPNETKTYFWKVQHHMPTKDEPCKAWAYFSD 979
Qy 947 VDLEKDVHSGLIGPLLVCHTNTLNPAGROVTVQEPALFFFTIDETKSNVYFENNERCR 1006
Db 980 VDLEKDVHSGLIGPLLVCHTNTLNPAGROVTVQEPALFFFTIDETKSNVYFENNERCR 1039
Qy 1007 APCNQIMQEDPFPKYNRFAHNGYIMDTLPLGLVMAQDQIRIRWYLLSMGNSNENIHSIFSG 1066
Db 1040 APCNQIMQEDPFPKYNRFAHNGYIMDTLPLGLVMAQDQIRIRWYLLSMGNSNENIHSIFSG 1099
Qy 1067 HVFTVRKKEEYKMAIYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNK 1126
Db 1100 HVFTVRKKEEYKMAIYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNK 1159
Qy 1127 COTPLGMAASHIRDFQITASQYQGWAPKLAHLHYSGSINAWSTKEPFSWIKVDLLPMI 1186
Db 1160 COTPLGMAASHIRDFQITASQYQGWAPKLAHLHYSGSINAWSTKEPFSWIKVDLLPMI 1219
Qy 1187 IHGKITQAGARQFSSLYISQFIIMYSLDGKKWOTYRGNSTGTLMVFFGNVDSGKKNIF 1246
Db 1220 IHGKITQAGARQFSSLYISQFIIMYSLDGKKWOTYRGNSTGTLMVFFGNVDSGKKNIF 1279
Qy 1247 NPPIIARVIRLHPHYSTRSTURMELMGCDLNSCMPLGMSKAI SDAQITASSYFTNMF 1306
Db 1280 NPPIIARVIRLHPHYSTRSTURMELMGCDLNSCMPLGMSKAI SDAQITASSYFTNMF 1339
Qy 1307 ATWSPSKARLHLQGRSNARPOVNNPKWLQYDFQKTMKVTGVTGQVKSLTSMYVKEF 1366
Db 1340 ATWSPSKARLHLQGRSNARPOVNNPKWLQYDFQKTMKVTGVTGQVKSLTSMYVKEF 1399
Qy 1367 LISSQDGHQWTLFPQNGKVKVQFQNGQDSFTPVNSLDPPLLTRYLRIHPQSWHQAIALR 1426
Db 1400 LISSQDGHQWTLFPQNGKVKVQFQNGQDSFTPVNSLDPPLLTRYLRIHPQSWHQAIALR 1459
Qy 1427 MEVLGCEAQDLY 1438
Db 1460 MEVLGCEAQDLY 1471

RESULT 7

AAR12971
ID AAR12971 standard; protein; 1440 AA.

XX AAR12971;

XX 25-MAR-2003 (updated)

DT 09-JAN-2003 (updated)

DT 02-OCT-1991 (first entry)
XX Factor VIII:SQ.
XX Factor VIII; B domain; haemophilia.
XX Unidentified.
XX MO9109122-A.
XX 27-JUN-1991.
XX 06-DEC-1990; 90WO-SE00809.
XX 15-DEC-1989; 89SE-0004239.
XX (KABI) KABIVITRUM AB.
XX Almstedt AB, Hellstrom EM, Larsson K, Lind P, Sandberg HI;
XX Spira J, Sydowback MM;
XX WPI; 1991-208148/28.
XX Recombinant human factor VIII deriv. deoxyribonucleic acid -
XX encoding protein comprising two chains linked by segment of B domain.
XX Disclosure; Fig 1; 35pp; English.
XX The protein is a fusion between Phe 742 and Ser 1637 of the factor
XX VIII protein (factor VIII:SQ). In order to produce a
XX factor VIII deletion derivative that can be produced in vivo and/or
XX in vitro, to a two chain protein consisting of polypeptide chains of
XX 90 kD and 80 kD, the amino acid sequences surrounding Arg 740 and
XX Arg 1648 have to be conserved in order to preserve the structural
XX requirements for correct cleavage. In this example, amino acids 743
XX to 1636 of the full-length factor VIII polypeptide are deleted. A
XX new polypeptide chain is obtd. where there are 14 amino acids
XX linking Arg 740 and Arg 1648. Of these 14 amino acids, the sequence
XX of the five N-terminal ones directly corresponds to the five amino
XX acids following Arg 740 in full-length factor VIII. Also, the sequence
XX of the 12 C-terminal amino acids of the above 14 amino acids fragment
XX directly corresponds to the 12 amino acids preceding Glu 1689 in full-
XX length factor VIII, thus creating a 3 amino acid overlap between the
XX N- and C-terminal regions of the B-domain.
XX The factor VIII deriv. is useful for treating haemophilia or
XX haemophilia A. It has the biological characteristics of plasma derived
XX factor VIII.
XX In order to index this example, the factor VIII:QD amino acid
XX sequence was retrieved from W0800831 (AAP80265).
XX The amino acid numbering in the above comments is reproduced from the
XX fig. description in the specification. Note that Arg 740 is Arg 742
XX in the indexed sequence, etc. Also, Asn 745 (N-terminal link overlap)
XX is Asp 747 in AAP80265, but indexed as Asn to reproduce the fusion
XX fragment as shown in fig.1, and Asn 745 (C-terminal link overlap).
XX (Updated on 09-JAN-2003 to add missing OS field.)
XX (Updated on 25-MAR-2003 to correct PI field.)
XX Sequence 1440 AA;

Query Match 99.6%; Score 7659.5; DB 12; Length 1440;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1434; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 1 ATRRYLGAVELSWDMQSDLGELPVDARPPRPYKSPFPNTSVVYKKTLEFVTHLEN 60
Db 2 ATRRYLGAVELSWDMQSDLGELPVDARPPRPYKSPFPNTSVVYKKTLEFVTHLEN 61
Qy 61 IAKPRPPMGLGPTIQAEVYDVTVTILKNWASHPVSILHAYGVSWKASEGAYDDQTSQ 120
Db 62 IAKPRPPMGLGPTIQAEVYDVTVTILKNWASHPVSILHAYGVSWKASEGAYDDQTSQ 121
Qy 121 REKEDDKVFPQSGSHYVQVLKENGPMASDPLCTYTSYLSHVDLVKDLNSGLIGALIVCR 180

Db 122 REKEDDKVPFGSGSHYVQVLEKNGPMASDPLCLITYSLYSHVDLVKDLNSGLIGALLVCR 181
QY 181 EGS�AKEKTQTLHKFILLFAVPFDEGKSWHSETKNSLMQDRDAASARAWPMQHTVNGVNR 240
Db 182 EGS�AKEKTHTLHKFILLFAVPFDEGKSWHSETKNSLMQDRDAASARAWPMQHTVNGVNR 241
QY 241 S-LPGLIGCHRKSVYWHVIGMGTTPPEVHSIFLGHTFLVRNHRQASLEISPIIFLTAQTL 299
Db 242 SLAPGLIGCHRKSVYWHVIGMGTTPPEVHSIFLGHTFLVRNHRQASLEISPIIFLTAQTL 301
QY 300 LMDLGGFLFCHLSSHQHDGMEAYVKVDSCEPEPQRLRMKNNEAEYDDDLTDTSEMDVVR 359
Db 302 LMDLGGFLFCHLSSHQHDGMEAYVKVDSCEPEPQRLRMKNNEAEYDDDLTDTSEMDLVR 361
QY 360 FDDNDSPSFIQIRSAVAKHKPTWHYIAAEEEDWDYAPLVLA PDDRSYKSQYLNNGPQRI 419
Db 362 FDDNDSPSFIQIRSAVAKHKPTWHYIAAEEEDWDYAPLVLA PDDRSYKSQYLNNGPQRI 421
QY 420 GRKYKVRFMAYTDTFTKTRATQHESGILGPLLYGEVGDITLLIIFNQASRPYNIYPHG 479
Db 422 GRKYKVRFMAYTDTFTKTRATQHESGILGPLLYGEVGDITLLIIFNQASRPYNIYPHG 481
QY 480 ITDVRPLYSRRLPGKVKHLDPILPGEIPKYKWTVTVEDGPTKSPRCILTRYVSSFVNM 539
Db 482 ITDVRPLYSRRLPGKVKHLDPILPGEIPKYKWTVTVEDGPTKSPRCILTRYVSSFVNM 541
QY 540 ERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSMYLTENIQRFLEPNA 599
Db 542 ERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSMYLTENIQRFLEPNA 601
QY 600 GVQLEDPFOASNIMHSINGVYFDSLSVCLHEVAYWYILSTGAQTDFLSVFSGYTFK 659
Db 602 GVQLEDPFOASNIMHSINGVYFDSLSVCLHEVAYWYILSTGAQTDFLSVFSGYTFK 661
QY 660 HKMAYEDTLTLFPSETVMSMENGLWILGCHNSDFRNGMTALLKYSSCKNTGDYY 719
Db 662 HKMAYEDTLTLFPSETVMSMENGLWILGCHNSDFRNGMTALLKYSSCKNTGDYY 721
QY 720 EDSYEDIAYLLSKNNAIEPRSPQNPVLKRQREITRTTLQSDQEEIDYDITISVEMK 779
Db 722 EDSYEDIAYLLSKNNAIEPRSPQNPVLKRQREITRTTLQSDQEEIDYDITISVEMK 781
QY 780 KEDFDIYDEENQSPRSFQKTRHYFTAAVERLWDYGMSSSPHVLNRQAQSGVPOFKKY 839
Db 782 KEDFDIYDEENQSPRSFQKTRHYFTAAVERLWDYGMSSSPHVLNRQAQSGVPOFKKY 841
QY 840 VFQEFYDGSFTQPLYRGEINEHLGILGPYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYE 899
Db 842 VFQEFYDGSFTQPLYRGEINEHLGILGPYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYE 901
QY 900 EDORQAEPRKPNFVKNETKTYFWKQHHMPTKDEFDCKAWAYFSDVDLEKDVHSGLIG 959
Db 902 EDORQAEPRKPNFVKNETKTYFWKQHHMPTKDEFDCKAWAYFSDVDLEKDVHSGLIG 961
QY 960 PLLVCHTNTLPAHGQVTVQEPALFPTTIPDETQSWYFTENMERNCRAPCNIQMEDPTFK 1019
Db 962 PLLVCHTNTLPAHGQVTVQEPALFPTTIPDETQSWYFTENMERNCRAPCNIQMEDPTFK 1021
QY 1020 ENTRFHAINGYIMDTLPLVMAQDRIWYLLSMGNSNENIHSIFPSGHVTVTRKGEYKM 1079
Db 1022 ENTRFHAINGYIMDTLPLVMAQDRIWYLLSMGNSNENIHSIFPSGHVTVTRKGEYKM 1081
QY 1080 ALYNLPVGVETVEMLPSPKAGIWRVECLIGEHLHAGMSTLFLVYSNKKOPLGMASGHIR 1139
Db 1082 ALYNLPVGVETVEMLPSPKAGIWRVECLIGEHLHAGMSTLFLVYSNKKOPLGMASGHIR 1141
QY 1140 DFOITASGOYGQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQGARQKF 1199
Db 1142 DFOITASGOYGQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQGARQKF 1201
QY 1200 SSLYISQFIIMYSLDGKWKQTYRGNSTGTLWVFGNVDSGSIKHNIFNPPIIARYIRLHP 1259
Db 1202 SSLYISQFIIMYSLDGKWKQTYRGNSTGTLWVFGNVDSGSIKHNIFNPPIIARYIRLHP 1261

QY 1260 THYSIRSTRMLMGLMCDLNSCSMPLGWSKASDAQITASSYFTNNMFATWSPSKARLHLQ 1319
Db 1262 THYSIRSTRMLMGLMCDLNSCSMPLGWSKASDAQITASSYFTNNMFATWSPSKARLHLQ 1321
QY 1320 GRSNARVPQNNPKWELQVDFOKTKWKTGVTGTQGVKSLLTSMYVKEFLISSODGHQWTL 1379
Db 1322 GRSNARVPQNNPKWELQVDFOKTKWKTGVTGTQGVKSLLTSMYVKEFLISSODGHQWTL 1381
QY 1380 FFONGKVKVQGNQDSFTPVVNSLDPPLLTRILRIHPQSWHQIALRMEVLGCEAODLY 1438
Db 1382 FFONGKVKVQGNQDSFTPVVNSLDPPLLTRILRIHPQSWHQIALRMEVLGCEAODLY 1440

RESULT 8
AAE10832
ID AAE10832 standard; Protein; 1459 AA.
XX
AC AAE10832;
XX
DT 18-DEC-2001 (first entry)
XX Human factor VIII mutein encoded by vector pTGF8-2hyg-s.
XX Human; haemostatic; coagulant; blood clotting factor; factor VIII;
XX factor IX; therapy; haemophilia A; mutant; mutein.
XX Homo sapiens.
OS
OS Synthetic.
XX
FH Key
FT Peptide 1..19
FT Protein /label= Signal_peptide
FT /product= "Human mature factor VIII mutein"
FT Region 750..775
FT /label= Linker-peptide
XX
PN W0200170968-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-EP03220.
XX
PR 22-MAR-2000; 2000EP-0106225.
PR 08-MAY-2000; 2000US-203249P.
XX
PA (OCTA-) OCTAGENE GMBH.
XX
PI Hauser C, Hoerster A, Schroeder C, Lehnerer M;
XX
DR WPI; 2001-590175/66.
DR N-PSDB; AAD18176.
XX
PT Recombinantly producing human blood coagulation factors VIII and IX for
XX use in treating hemophilia -
PS Claim 17; Page 86-90; 104pp; English.
XX
CC The present invention relates to an improved method for the production
CC of recombinant human blood clotting factors, especially factors VIII and
CC IX, using an immortalised cell line stably expressing viral transcription
CC activators and carrying a vector comprising a promoter and a sequence
CC encoding the blood coagulation factor. The factor VIII mutein or a gene
CC transfer vector is used in the preparation of agents for treating
CC haemophilia, especially haemophilia A. The present sequence is
CC human factor VIII mutein encoded by vector pTGF8-2hyg-s. The vector
CC contains a silent mutation, resulting in a factor VIII mutein having
CC the substitution of the B-domain of wild-type factor VIII by the
XX linker peptide.
SQ Sequence 1459 AA;

Query Match				99.2%;	Score 7626;	DB 22;	Length 1459;
Best Local Similarity				99.1%;	Pred. No. 0;		
Matches 1431;				Conservative	0;	Mismatches	3;
						Indels	10;
						Gaps	2;
QY	1	ATRRYYLGAVELSWDMOSDLGELPVDARPPRPVKSPFFNTSVYKKTLPVFETHLPN	60				
DB	20	ATRRYYLGAVELSWDMOSDLGELPVDARPPRPVKSPFFNTSVYKKTLPVFETHLPN	79				
QY	61	IAKPRPPWMLGPTTQAEVYDVTVITLKNMASHPVSLHAGVSVYKASEGAEYDDQTSQ	120				
DB	80	IAKPRPPWMLGPTTQAEVYDVTVITLKNMASHPVSLHAGVSVYKASEGAEYDDQTSQ	139				
QY	121	REKEDDKVPFGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR	180				
DB	140	REKEDDKVPFGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR	199				
QY	181	EGSLAKEKTQTLHKETILLFAVDEGKSWHSETKNSLMQDRDAASARAPKMHVTVNGVYNR	240				
DB	200	EGSLAKEKTQTLHKETILLFAVDEGKSWHSETKNSLMQDRDAASARAPKMHVTVNGVYNR	259				
QY	241	SLPGLIGCHRSVYVHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTLL	300				
DB	260	SLPGLIGCHRSVYVHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTLL	319				
QY	301	MDLGQFLFCHTSSHQHDGMEAYVVDSCPEEPOLRMKNNEEAEDYDDDLTDSMDVYRF	360				
DB	320	MDLGQFLFCHTSSHQHDGMEAYVVDSCPEEPOLRMKNNEEAEDYDDDLTDSMDVYRF	379				
QY	361	DDNSPSFIQIRSVAKGPKTWVHYIAAEEEDWDYAPLVAPDDRSYKSOYLNNGPQRIQ	420				
DB	380	DDNSPSFIQIRSVAKGPKTWVHYIAAEEEDWDYAPLVAPDDRSYKSOYLNNGPQRIQ	439				
QY	421	RKYKVRFMAYTDETFKTREAIQHESGILGPLYGEVGDITLLIIFKNQASRPNTIYPHGI	480				
DB	440	RKYKVRFMAYTDETFKTREAIQHESGILGPLYGEVGDITLLIIFKNQASRPNTIYPHGI	499				
QY	481	TDVRLPYRRLPKGVKHLKDFILPGEIFKYKMTVTVEDGPTKSDPCLTRYYSFVNM	540				
DB	500	TDVRLPYRRLPKGVKHLKDFILPGEIFKYKMTVTVEDGPTKSDPCLTRYYSFVNM	559				
QY	541	RLIASGLIGPLLCYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIORPLNPAG	600				
DB	560	RLIASGLIGPLLCYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIORPLNPAG	619				
QY	601	VQLEDPFOASNIMHSINGVYVDSLSQSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH	660				
DB	620	VQLEDPFOASNIMHSINGVYVDSLSQSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH	679				
QY	661	KWYVEDTTLTLPFSGETVFMSENPGILWILGCHNSDFRNKGMTALLKVSSCDKNTGDIYE	720				
DB	680	KWYVEDTTLTLPFSGETVFMSENPGILWILGCHNSDFRNKGMTALLKVSSCDKNTGDIYE	739				
QY	721	DSYEDISAYLLSKNNAIEPRSPQNPVVKRQR-----EITRTTTLQSDQEEIDYDDTI	774				
DB	740	DSYEDISAYLLSKNNAIEPRSPQN-----SRHQAYRYRGEITRTTTLQSDQEEIDYDDTI	795				
QY	775	SVEMKKEPFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRAQSGSVP	834				
DB	796	SVEMKKEPFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRAQSGSVP	855				
QY	835	QFKVVFQEFQFTGSGTQPLRYGELNEHLGLLPYIRAEVEDNIMVTFRNQASRPYSFYSS	894				
DB	856	QFKVVFQEFQFTGSGTQPLRYGELNEHLGLLPYIRAEVEDNIMVTFRNQASRPYSFYSS	915				
QY	895	LISYEDORQGAEPKRNFKVNETKTYFMKVQHMAPTKDQFDCKAWAYFSDVDLEKDVH	954				
DB	916	LISYEDORQGAEPKRNFKVNETKTYFMKVQHMAPTKDQFDCKAWAYFSDVDLEKDVH	975				
QY	955	SGLIGPLLYCHTNTLNPAHGRQVTVQSFALFTTIFDETSKSWYFTENMERNCRAPCNIQME	1014				
DB	976	SGLIGPLLYCHTNTLNPAHGRQVTVQSFALFTTIFDETSKSWYFTENMERNCRAPCNIQME	1035				
QY	1015	DPTFKENYRFAHNGYIMDTLPGLVMAQDORIRWYLLSMGSNENIHSIHFSGHVFTVRKK	1074				

DB	1036	DPTFKENYRFAHNGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIHFSGHVFTVRKK	1095
QY	1075	EYKXALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGMA	1134
DB	1096	EYKXALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGMA	1155
QY	1135	SGHIRDFOITASGOYQOWAPKLAHLHYSGSINASTKEPFSWIKVDLLAPMIHGIKTQG	1194
DB	1156	SGHIRDFOITASGOYQOWAPKLAHLHYSGSINASTKEPFSWIKVDLLAPMIHGIKTQG	1215
QY	1195	ARQFSSLYISQFIIMYSLDGKKQTVRGNSGTGLMVFFGNVDSSGIGHNIFNPPILIARY	1254
DB	1216	ARQFSSLYISQFIIMYSLDGKKQTVRGNSGTGLMVFFGNVDSSGIGHNIFNPPILIARY	1275
QY	1255	IRLHPTHYSIRSTRMELMGCNLNSCMPLGMSKASDAQITASSYFTNMFATWSPSKA	1314
DB	1276	IRLHPTHYSIRSTRMELMGCNLNSCMPLGMSKASDAQITASSYFTNMFATWSPSKA	1335
QY	1315	RLHLQGRSNARPOVNNPKEWLQVDFQKTMKVTVGTQGVKSLLTSMYVKEFLISSQDG	1374
DB	1336	RLHLQGRSNARPOVNNPKEWLQVDFQKTMKVTVGTQGVKSLLTSMYVKEFLISSQDG	1395
QY	1375	HOWTLFFQNGVKVYFQGNQDSFTPVNSLDPELLTRYLRHPQSWHQAIALRMEVLGCEA	1434
DB	1396	HOWTLFFQNGVKVYFQGNQDSFTPVNSLDPELLTRYLRHPQSWHQAIALRMEVLGCEA	1455
QY	1435	QDLY 1438	
DB	1456	QDLY 1459	
RESULT 9			
AAE10833			
ID	AAE10833	standard; Protein; 1459 AA.	
XX	AAE10833;		
XX	18-DEC-2001	(first entry)	
XX	Human factor VIII mutein encoded by vector pTGF8-3.		
XX	Human; haemostatic; coagulant; blood clotting factor; factor VIII;		
KW	factor IX; therapy; haemophilia A; mutant; mutein.		
OS	Homo sapiens.		
OS	Synthetic.		
Key	Location/Qualifiers		
FT	Peptide	1..19	
FT	Protein	/label= Signal peptide	
FT	Misc-difference	20..1459	
FT	Region	/product= "Human mature factor VIII mutein"	
FT	Misc-difference	181	
FT	Region	/note= "Wild type Val substituted with Ala"	
FT	Misc-difference	760..775	
FT	Misc-difference	1350	
FT	Misc-difference	/note= "Wild type Val substituted with Glu"	
XX	WO200170968-A2.		
XX	27-SEP-2001.		
XX	21-MAR-2001;	2001WO-EP03220.	
XX	22-MAR-2000;	2000EP-0106225.	
XX	08-MAY-2000;	2000US-203249E.	
XX	(OCTA-) OCTAGENE GMBH.		
PI	Hauser C, Hoerster A, Schroeder C, Lehnerer M;		
XX			

DR WPI; 2001-590175/66.
XX N-PSDB; AAD181176.
PT Recombinantly producing human blood coagulation factors VIII and IX for
PS use in treating hemophilia -
XX Claim 17; Page 99-103; 104pp; English.
XX
CC The present invention relates to an improved method for the production
CC of recombinant human blood clotting factors, especially factors VIII and
CC IX, using an immortalised cell line stably expressing viral transcription
CC activators and carrying a vector comprising a promoter and a sequence
CC encoding the blood coagulation factor. The factor VIII mutin or a gene
CC transfer vector is used in the preparation of agents for treating
CC haemophilia, especially haemophilia A. The present sequence is human
CC factor VIII mutin encoded by vector pTGF8-3. This sequence contains
CC 2 mutations and the B-domain of wild-type factor VIII is substituted
CC with a linker peptide.
XX
SQ Sequence 1459 AA;

Query Match 99.0%; Score 7616; DB 22; Length 1459;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1429; Conservative 0; Mismatches 5; Indels 10; Gaps 2;

QY 1 ATRRYILGAVELSDYMQSDLGELPVDAREPPRPVPSKSPFNTSVYKTLFVEFTVHLEN 60
DB 20 ATRRYILGAVELSDYMQSDLGELPVDAREPPRPVPSKSPFNTSVYKTLFVEFTVHLEN 79
QY 61 IAKPRPPWMLLOPTQAEVYDVTVTILKMAHPVSLHAGVSYKASGAYDDOTSQ 120
DB 80 IAKPRPPWMLLOPTQAEVYDVTVTILKMAHPVSLHAGVSYKASGAYDDOTSQ 139
QY 121 REKEDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
DB 140 REKEDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 199
QY 181 EGSIAKEKTLHLKFIILFAVDEGKSWHSETKNSLMQDRDAASARAPKQHTVNGVNR 240
DB 200 EGSIAKEKTLHLKFIILFAVDEGKSWHSETKNSLMQDRDAASARAPKQHTVNGVNR 259
QY 241 SLPLGIGCHRSYVHVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLL 300
DB 260 SLPLGIGCHRSYVHVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLL 319
QY 301 MDLGOFLFCHISSHQHDGMEAYVKVDSCEEPQLRMKNNEAEEDYDDDLTDSMDVRF 360
DB 320 MDLGOFLFCHISSHQHDGMEAYVKVDSCEEPQLRMKNNEAEEDYDDDLTDSMDVRF 379
QY 361 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPADDRSYKSYLNNGPORIG 420
DB 380 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPADDRSYKSYLNNGPORIG 439
QY 421 RYKVKVRFMAYTDFTFKTREAIOHESGILGPLLYGEVGDTLIIIFKNQASRPNIYPHGI 480
DB 440 RYKVKVRFMAYTDFTFKTREAIOHESGILGPLLYGEVGDTLIIIFKNQASRPNIYPHGI 499
QY 481 TDVRPLYSRRLPGVRLKDPFPLPGEI PKYKMTVTVEDEGTPKSDRCLTRYSSFNME 540
DB 500 TDVRPLYSRRLPGVRLKDPFPLPGEI PKYKMTVTVEDEGTPKSDRCLTRYSSFNME 559
QY 541 RDLASGLIGPLLCYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 600
DB 560 RDLASGLIGPLLCYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 619
QY 601 VQLEDPFQASNMHMSINGVYFDSLSQSVCLHEVAYWYIISIGAQTDFLSVFFSGYTFKH 660
DB 620 VQLEDPFQASNMHMSINGVYFDSLSQSVCLHEVAYWYIISIGAQTDFLSVFFSGYTFKH 679
QY 661 KNVYEDTLTLPPSGEYTFVMSMNPGLWILGCHNSDFRNKGMTALLKVSSCDKNTGYYE 720
DB 680 KNVYEDTLTLPPSGEYTFVMSMNPGLWILGCHNSDFRNKGMTALLKVSSCDKNTGYYE 739

QY 721 DSYEDISAYLLSKNNAIEPRSFSONPPVLKRHOR-----EITRTTLOSDEBIDYDDTI 774
DB 740 DSYEDISAYLLSKNNAIEPRSFSON-----SRHQVYRRGEITRTTLOSDEBIDYDDTI 795
QY 775 SVMKKEDFDIYDEDENQSPRGFQKTRHYFTAAVERLWDYGMSSSPHYLRNRAQSGSVP 834
DB 796 SVMKKEDFDIYDEDENQSPRGFQKTRHYFTAAVERLWDYGMSSSPHYLRNRAQSGSVP 855
QY 835 QFKKVVQFBFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTRNQASRPYFYSS 894
DB 856 QFKKVVQFBFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTRNQASRPYFYSS 915
QY 895 LISYEDQOQABPRKNFKVKNETKTYFKVQVHMAPTKDEDFCKAWAYFSDVDLEKDVH 954
DB 916 LISYEDQOQABPRKNFKVKNETKTYFKVQVHMAPTKDEDFCKAWAYFSDVDLEKDVH 975
QY 955 SGLIGPLLCHNTNLNPAHGRQVTVQEFALFTTIDETKSWYFTENWERNCRAPCNIOME 1014
DB 976 SGLIGPLLCHNTNLNPAHGRQVTVQEFALFTTIDETKSWYFTENWERNCRAPCNIOME 1035
QY 1015 DPTFKENYRPHALNGYIMDTLPCLVMAQDQRIRWYLLSMGNSNENIHSIHFSGHVFTVRKK 1074
DB 1036 DPTFKENYRPHALNGYIMDTLPCLVMAQDQRIRWYLLSMGNSNENIHSIHFSGHVFTVRKK 1095
QY 1075 EYKMALYNLYPGVFTVEMLPKAGIWRVVECLIGELHAGMSTLFLVYSNKCQTPLGMA 1134
DB 1096 EYKMALYNLYPGVFTVEMLPKAGIWRVVECLIGELHAGMSTLFLVYSNKCQTPLGMA 1155
QY 1135 SGHIRDFQITASGOYQONAPKLARLHYSINAWSTKEPFSWKVDLLAPMIHGIKTQG 1194
DB 1156 SGHIRDFQITASGOYQONAPKLARLHYSINAWSTKEPFSWKVDLLAPMIHGIKTQG 1215
QY 1195 ARQKFSLSYISQFIIMYSLDGKKQWYRGNSGTGLMVFFGNVDSSGINKNIFNPPIIARY 1254
DB 1216 ARQKFSLSYISQFIIMYSLDGKKQWYRGNSGTGLMVFFGNVDSSGINKNIFNPPIIARY 1275
QY 1255 IRLHPTHYSIRSTRLMELMGCDLNSCMLPLGMSKASISDAQITASSYFTNMFWATWSPSKA 1314
DB 1276 IRLHPTHYSIRSTRLMELMGCDLNSCMLPLGMSKASISDAQITASSYFTNMFWATWSPSKA 1335
QY 1315 RLHLQGRSNARPOVNNPKWLQVDFQKTMKVTVGTQGVKSLTSMYKKEFLISSQDG 1374
DB 1336 RLHLQGRSNARPOVNNPKWLQVDFQKTMKVTVGTQGVKSLTSMYKKEFLISSQDG 1395
QY 1375 HQTLLFFQNGKVKVFOGQNDSTFPVNVNSLDPPLLTRYLRIHPOSWVHOIALRMEVLGCEA 1434
DB 1396 HQTLLFFQNGKVKVFOGQNDSTFPVNVNSLDPPLLTRYLRIHPOSWVHOIALRMEVLGCEA 1455
QY 1435 QDLY 1438
DB 1456 QDLY 1459
RESULT 10
AAE10827
ID AAE10827 standard; Protein; 1459 AA.
XX
XX AAE10827;
XX
XX 18-DEC-2001 (first entry)
DT
XX Human factor VIII mutin encoded by vector pTGF8-1.
XX
XX Human; haemostatic; coagulant; blood clotting factor; factor VIII;
KW factor IX; therapy; haemophilia A; mutant; muten.
XX Homo sapiens.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Peptide 1..19
FT /label= signal_peptide
FT Protein 20..1459

AAR80265
 ID AAP80265 standard; protein; 1516 AA.
 AC AAP80265;
 XX 25-MAR-2003 (updated)
 DT 10-OCT-1990 (first entry)
 XX
 DE Modified factor VIII:C sequence with the Q744-D1563 deletion.
 XX
 KW Modified factor VIII:C; maturation polypeptide; haemophilia;
 KM blood coagulation; QD deletion.
 XX
 OS Homo sapiens.
 XX
 PN MO8800831-A.
 XX
 PD 11-FEB-1988.
 XX
 PF 31-JUL-1987; 87WO-US01814.
 XX
 PR 01-AUG-1986; 86US-0893375.
 XX
 PA (BIOJ) BIOGEN NV.
 PA (PASE/) PASEK M P.
 XX
 PI Pasek MP;
 XX
 DR WPI; 1988-049866/07.
 DR N-PSDB; AAN80444.
 XX
 PT New DNA sequences encoding modified factor VIII:C - with deletion of DNA
 PT encoding maturation polypeptide, useful for high yield transformation.
 XX
 PS Claim 3; Page 51-52-53-54; 97pp; English.
 XX
 CC A major part of the sequence encoding the maturation polypeptide of
 CC factor VIII:C is deleted, i.e. Gln 744 - Asp 1563. The QD deletion
 CC retains approximately 90 amino acids of the maturation polypeptide
 CC (four amino acids at the N-terminal end and 86 amino acids at
 CC the C-terminal end). The full length Factor VIII:C cDNA has two
 CC changes with respect to the published sequence (EPO application 160457):
 CC Cys to Cys at Leu 242 and TTC to CTC change at amino acid residue 1880
 CC (Phe to Leu).
 CC The product is produced in approx. 20 times higher
 CC yields than previous recombinant produced factor VIII:C and are more
 CC easily purified. The peptide is used for treating haemophilia A, both
 CC acute and prolonged bleeding.
 CC See also AAN80446 and AAN80447.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX
 SQ Sequence 1516 AA;
 Query Match 99.08; Score 7611.5; DB 9; Length 1516;
 Best Local Similarity 94.74; Pred. No. 0;
 Matches 1434; Conservative 1; Mismatches 3; Indels 77; Gaps 2;
 QY 1 ATRRYLGAVELSDVYQSDLGELPVDARPPRPVPSFPNTSVVYKTLFVBEFTVHLFN 60
 DB 2 ATRRYLGAVELSDVYQSDLGELPVDARPPRPVPSFPNTSVVYKTLFVBEFTVHLFN 61
 QY 61 IAKPRPPWMLGPTTQAEVYDVTVTITLKNASHPVSLHGVSYNWKASBGAEDDQTSQ 120
 DB 62 IAKPRPPWMLGPTTQAEVYDVTVTITLKNASHPVSLHGVSYNWKASBGAEDDQTSQ 121
 QY 121 REKEDDKVFGSGSYTVQVNLKENGPMASDPLCLTYSYLGHVDLVKDLNSGLIGALLVCR 180
 DB 122 REKEDDKVFGSGSYTVQVNLKENGPMASDPLCLTYSYLGHVDLVKDLNSGLIGALLVCR 181
 QY 181 EGSIAKSKTQTLHKFILLFAVFDGSKSWHSETKNSLMQDRDAASARAWPKMHTVGVYNR 240
 DB 182 EGSIAKSKTQTLHKFILLFAVFDGSKSWHSETKNSLMQDRDAASARAWPKMHTVGVYNR 241

QY 241 S-LPLGLIGCHRSVYWHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTL 299
 DB 242 SLPLGLIGCHRSVYWHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTL 301
 QY 300 LMDLGQFLFCHISSHQHDMAYVKVDSCEPQPOLRMKNNEAEADYDDDLTDTSEMDVVR 359
 DB 302 LMDLGQFLFCHISSHQHDMAYVKVDSCEPQPOLRMKNNEAEADYDDDLTDTSEMDLVR 361
 QY 360 FDDNSPFIQIRSVAKGHPKTVWHVYIAAEEEDWDYAPLVAPDDRSYKSYQYLNNGPQRI 419
 DB 362 FDDNSPFIQIRSVAKGHPKTVWHVYIAAEEEDWDYAPLVAPDDRSYKSYQYLNNGPQRI 421
 QY 420 GRKYKKVRMAYTDETFKTRAIQHESGILGPLLYGEGVDTLLIIFKNQASRYNIYPHG 479
 DB 422 GRKYKKVRMAYTDETFKTRAIQHESGILGPLLYGEGVDTLLIIFKNQASRYNIYPHG 481
 QY 480 ITDVRPLYSRRLPKGVKHLKDPFILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSFVNM 539
 DB 482 ITDVRPLYSRRLPKGVKHLKDPFILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSFVNM 541
 QY 540 ERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRFNP 599
 DB 542 ERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRFNP 601
 QY 600 GVQLEDPEFOASNIMHSINGYVFDLSQLSVCLHEVAYWYLSIGAQDTFLSVFSGYTFK 659
 DB 602 GVQLEDPEFOASNIMHSINGYVFDLSQLSVCLHEVAYWYLSIGAQDTFLSVFSGYTFK 661
 QY 660 HKMVEDTLTLPPFSGETVFMSENPLGTLGCHNSDFRNRGMTALLKVVSSCDKNTGDDY 719
 DB 662 HKMVEDTLTLPPFSGETVFMSENPLGTLGCHNSDFRNRGMTALLKVVSSCDKNTGDDY 721
 QY 720 EDSYEDISAYLLSKNNAIBRSPF----- 742
 DB 722 EDSYEDISAYLLSKNNAIBRSPSQDPLAWNHYGTQIPKEEWSQESPEKTAFFKKD 781
 QY 743 -----SONPVLRKHOREITRITLQ 763
 DB 782 ILSINACESHAIATAINEGQNKPEIEVTWAKQGRTERLCSQNPVPLKRHOREITRITLQ 841
 QY 764 DQBEIDYDDTISVEMKEDPDIDYDENOSPRSQKTRHYFTAAVERLDYGMSSSPHV 823
 DB 842 DQBEIDYDDTISVEMKEDPDIDYDENOSPRSQKTRHYFTAAVERLDYGMSSSPHV 901
 QY 824 LRNRAQSGSVQPKVVFQFTDGSFTQPLRYGELNEHLGLGPIYAEVEDNIMVTFRN 883
 DB 902 LRNRAQSGSVQPKVVFQFTDGSFTQPLRYGELNEHLGLGPIYAEVEDNIMVTFRN 961
 QY 884 QASRPYSFYSSLSIYSEEDQOQAEPRKNFVKPNETKTYFMKVOHMAPTKDEPDCCKAWAY 943
 DB 962 QASRPYSFYSSLSIYSEEDQOQAEPRKNFVKPNETKTYFMKVOHMAPTKDEPDCCKAWAY 1021
 QY 944 FSDVLEKDVHSLIGPLLVCHTNTLNPAHGRVTVQEPALPFTIIDEKTSWTFENNER 1003
 DB 1022 FSDVLEKDVHSLIGPLLVCHTNTLNPAHGRVTVQEPALPFTIIDEKTSWTFENNER 1081
 QY 1004 NCRAPCNIMQEDPTFKENYRFAINGYIMDTLPLGVMAQDQIRWYLLSGNSNENIHSIH 1063
 DB 1082 NCRAPCNIMQEDPTFKENYRFAINGYIMDTLPLGVMAQDQIRWYLLSGNSNENIHSIH 1141
 QY 1064 FSGHVTFRKKEBYKVALYNLYPGVPEVEMLPKAGIWRVECLIGELHAGNSTFLVY 1123
 DB 1142 FSGHVTFRKKEBYKVALYNLYPGVPEVEMLPKAGIWRVECLIGELHAGNSTFLVY 1201
 QY 1124 SNKCQPLGNASGHIEDFQITASGOYGQWAPKLARLHYSGSINAWSTKEPFSIKVDLLA 1183
 DB 1202 SNKCQPLGNASGHIEDFQITASGOYGQWAPKLARLHYSGSINAWSTKEPFSIKVDLLA 1261
 QY 1184 PMIHGKITQARQKFSLSYISQFIWYSLDGKKWQYRGNSTGLMVFNGVDSGSIKH 1243
 DB 1262 PMIHGKITQARQKFSLSYISQFIWYSLDGKKWQYRGNSTGLMVFNGVDSGSIKH 1321
 QY 1244 NIFNPPIIARYIRLHPHYSIRSTRMLMELMGCDLNSCSMPLGMESKAISDAQITASSYFT 1303

Db 1322 NIFNPPIIARIIRLHTYSIRSLRMELMGCDLNSCSMPLEMSKALSDAQITASSYFT 1381
Qy 1304 NMFATWSPSKARLHLQGRSNAWRPQVNNPKFQWLDVDFQRTMKVGTGVTGQVKSLTSMYV 1363
Db 1382 NMFATWSPSKARLHLQGRSNAWRPQVNNPKFQWLDVDFQRTMKVGTGVTGQVKSLTSMYV 1441
Qy 1364 KEFLISSQDGHQWTLFFQNGKVKVFOGQNSDFTFVNSLDPPLLTRYLRIRHPOSWVHQI 1423
Db 1442 KEFLISSQDGHQWTLFFQNGKVKVFOGQNSDFTFVNSLDPPLLTRYLRIRHPOSWVHQI 1501
Qy 1424 ALRMEVLGCEAODLY 1438
Db 1502 ALRMEVLGCEAODLY 1516

RESULT 12

ID AAB48842
ID AAB48842 standard; protein; 1424 AA.

XX AAB48842;

XX 13-MAR-2001 (first entry)

XX Mutant mature human factor VIII, SEQ ID NO:5.

XX Factor VIII; human; B domain; LRP-mediated plasma clearance;
KW receptor-dependent clearance; receptor-independent clearance;
KW half-life; haemophilia; mutant; mutein.

XX Homo sapiens.

XX WO200071714-A2.

XX 30-NOV-2000.

XX 24-MAY-2000; 2000WO-US14111.

XX 24-MAY-1999; 98US-0135847.

XX (AMNA-) AMERICAN NAT RED CROSS.

XX Saenko EL, Strickland DK;

XX WPI; 2001-025163/03.

XX Factor VIII mutants having increased half-life useful for treating
PT hemophilia, comprise one or more amino acid substitutions in the A2
PT and/or C2 domain of factor VIII -

PS Claim 9; Fig 2A-B; 121pp; English.

XX The invention relates to human factor VIII mutants comprising an amino
CC acid substitution at one or more positions in the A2 domain and/or an
CC amino acid substitution at one or more positions in the C2 domain.
CC The invention also encompasses a factor VIII mutant which lacks a B
CC domain (AAB48842). The factor VIII mutants have an increased half-life
CC in the bloodstream. The A2 domain mutants exhibit reduced LRP-dependent
CC (receptor-dependent) clearance of factor VIII, while C2 domain mutants
CC have reduced receptor-independent clearance. The invention also relates
CC to a method of using RAP (receptor associated protein), a protein which
CC inhibits LRP (low density lipoprotein related protein)-mediated ligand
CC internalisation, to increase the half-life of factor VIII. The mutant
CC factor VIII proteins, and nucleotides encoding them, are useful
CC for treating haemophilia. RAP, LRP-binding RAP mutants or fragments, and
CC nucleic acids encoding them may also be used in the treatment of
CC haemophilia, in combination with a mutant factor VIII protein or DNA of
CC the invention. The invention provides means of increasing the half-life
CC of factor VIII by reducing its clearance from plasma. The present
CC sequence represents a mutant mature human factor VIII which lacks a B
CC domain.

XX Sequence 1424 AA;

Qy 1322 NIFNPPIIARIIRLHTYSIRSLRMELMGCDLNSCSMPLEMSKALSDAQITASSYFT 1381
Db 1304 NMFATWSPSKARLHLQGRSNAWRPQVNNPKFQWLDVDFQRTMKVGTGVTGQVKSLTSMYV 1363
Qy 1382 NMFATWSPSKARLHLQGRSNAWRPQVNNPKFQWLDVDFQRTMKVGTGVTGQVKSLTSMYV 1441
Db 1364 KEFLISSQDGHQWTLFFQNGKVKVFOGQNSDFTFVNSLDPPLLTRYLRIRHPOSWVHQI 1423
Qy 1442 KEFLISSQDGHQWTLFFQNGKVKVFOGQNSDFTFVNSLDPPLLTRYLRIRHPOSWVHQI 1501
Db 1424 ALRMEVLGCEAODLY 1438
Qy 1502 ALRMEVLGCEAODLY 1516
Db 1502 ALRMEVLGCEAODLY 1516

Query Match 98.7%; Score 7592; DB 22; Length 1424;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1423; Conservative 0; Mismatches 1; Indels 14; Gaps 1;

Qy 1 ATRYYIYGAVELSDWDYQSDGLGELPVDARPPRPVKPFPPNTSVVYKKTLEFETVHLFN 60
Db 1 ATRYYIYGAVELSDWDYQSDGLGELPVDARPPRPVKPFPPNTSVVYKKTLEFETVHLFN 60
Qy 61 IAKPRPPMGLGPTIQAQVYDVTVITLKNASHPVSLHAGVSYWKASEGAEYDDQTSQ 120
Db 1 IAKPRPPMGLGPTIQAQVYDVTVITLKNASHPVSLHAGVSYWKASEGAEYDDQTSQ 120
Qy 121 REKEDDKVFPQGSHTYVQVLKENGPMASDPLCTYSYLSHVDLVKOLNSGLIGALLVCR 180
Db 121 REKEDDKVFPQGSHTYVQVLKENGPMASDPLCTYSYLSHVDLVKOLNSGLIGALLVCR 180
Qy 181 EGSIAKEKTQTLHKPFIILLFAVFDGKSWHSSTKNSLMQDRDAASARAWPKMHTVGVYNR 240
Db 181 EGSIAKEKTQTLHKPFIILLFAVFDGKSWHSSTKNSLMQDRDAASARAWPKMHTVGVYNR 240
Qy 241 SLPLGLIGCHRSVYVHVGIMGTTPPEVHSIFLEGHTFLVRNHRQASLSIPITFLTAQTLL 300
Db 241 SLPLGLIGCHRSVYVHVGIMGTTPPEVHSIFLEGHTFLVRNHRQASLSIPITFLTAQTLL 300
Qy 301 MDLQGLFLFCHISSHQDHMEAYVKVDSCEPQOLRMKNNEEADYDDDLTSEMDVVRP 360
Db 301 MDLQGLFLFCHISSHQDHMEAYVKVDSCEPQOLRMKNNEEADYDDDLTSEMDVVRP 360
Qy 361 DDNSPSFIQIRSVAKKHPKTMVHYIAAEEDWDYAPLVLAPDRSVKSYQVLNNGPQIG 420
Db 361 DDNSPSFIQIRSVAKKHPKTMVHYIAAEEDWDYAPLVLAPDRSVKSYQVLNNGPQIG 420
Qy 421 RYKVKRFMAYTDETFKTREAIQESGILGPLLYGEVGDITLLIIFKQASPYNIYPHGI 480
Db 421 RYKVKRFMAYTDETFKTREAIQESGILGPLLYGEVGDITLLIIFKQASPYNIYPHGI 480
Qy 481 TDVRPLYSRRLPKGVKHLKDPFIPGEIFKTKVTVVEDGPKSDPRLCTRYYSFVNE 540
Db 481 TDVRPLYSRRLPKGVKHLKDPFIPGEIFKTKVTVVEDGPKSDPRLCTRYYSFVNE 540
Qy 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKENVILFSVDENRSMYLTENORLEPNAG 600
Db 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKENVILFSVDENRSMYLTENORLEPNAG 600
Qy 601 VQLEDPEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660
Db 601 VQLEDPEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660
Qy 661 KMYVEDTLTLPFPGSETVFMSENPGWLILGCHNSDFNRGWTALLKVSSCDKXTGDTYE 720
Db 661 KMYVEDTLTLPFPGSETVFMSENPGWLILGCHNSDFNRGWTALLKVSSCDKXTGDTYE 720
Qy 721 DSYEDIISAYLLSKKNAIEPRSFQNPVVLKHHQREITRTTLQSDQERIDYDDTISVEMKK 780
Db 721 DSYEDIISAYLLSKKNAIEPRSFQNPVVLKHHQREITRTTLQSDQERIDYDDTISVEMKK 780
Qy 781 EDFDIYDEDNQSPRSFQKTRHYFIAAVERLWDYGMSSSPHYLRNRAQSGSVFQKXV 840
Db 781 EDFDIYDEDNQSPRSFQKTRHYFIAAVERLWDYGMSSSPHYLRNRAQSGSVFQKXV 840
Qy 841 FQETDGSFTQPLRYGELNEHLLGPGYIRAEVEDNIMVTFRNQASRPYSFSSLSYEE 900
Db 841 FQETDGSFTQPLRYGELNEHLLGPGYIRAEVEDNIMVTFRNQASRPYSFSSLSYEE 900
Qy 901 DORQABRKPNKVPKNETKTYFMKQHHMATKDFCKAWAYSDVDLEKDVHSLIGP 960
Db 901 DORQABRKPNKVPKNETKTYFMKQHHMATKDFCKAWAYSDVDLEKDVHSLIGP 960
Qy 961 LLVCHTNTLNPAHGRQVTVQEFALFFITFDSTKSWYFTENMERNCRAPCNIQMEDPTPK 1020
Db 961 LLVCHTNTLNPAHGRQVTVQEFALFFITFDSTKSWYFTENMERNCRAPCNIQMEDPTPK 1020
Qy 947 LLVCHTNTLNPAHGRQVTVQEFALFFITFDSTKSWYFTENMERNCRAPCNIQMEDPTPK 1006
Db 947 LLVCHTNTLNPAHGRQVTVQEFALFFITFDSTKSWYFTENMERNCRAPCNIQMEDPTPK 1006

QY 1021 NYRPHAINGYIMDTLPGLVMAQDQIRIRWYLLSMGSENENIHSIHPSGHVFTVRKKEBYMA 1080
 DB 1007 NYRPHAINGYIMDTLPGLVMAQDQIRIRWYLLSMGSENENIHSIHPSGHVFTVRKKEBYMA 1066
 QY 1081 LYNLYPGVFVEVEMLPKAGIRWVECLIGELHAGMSTFLVYVSNKQOTPLGMAASHIRD 1140
 DB 1067 LYNLYPGVFVEVEMLPKAGIRWVECLIGELHAGMSTFLVYVSNKQOTPLGMAASHIRD 1126
 QY 1141 FQITASQYQGWAPKPLARLHYSGSINAWSTKEPFSWIKVDLAPMIHGIKTQCARQKFS 1200
 DB 1127 FQITASQYQGWAPKPLARLHYSGSINAWSTKEPFSWIKVDLAPMIHGIKTQCARQKFS 1186
 QY 1201 SLVYSQFIIMYSLDGKQWYRGNSGTGLMVFFGNVDSSGKKNIFNPPIIARIYRLHPT 1260
 DB 1187 SLVYSQFIIMYSLDGKQWYRGNSGTGLMVFFGNVDSSGKKNIFNPPIIARIYRLHPT 1246
 QY 1261 HYSIRSTRMELMCDLNSCMPLGMSKATSDAQITASSYFTNMFATWSPSKARLHLOQ 1320
 DB 1247 HYSIRSTRMELMCDLNSCMPLGMSKATSDAQITASSYFTNMFATWSPSKARLHLOQ 1306
 QY 1321 RSNARWEPQVNNPKWLOVDFOKTMKVGTGVTGQVKSLLTSMYVKEFLISSQDGHQWTLF 1380
 DB 1307 RSNARWEPQVNNPKWLOVDFOKTMKVGTGVTGQVKSLLTSMYVKEFLISSQDGHQWTLF 1366
 QY 1381 FONGKVKVQGNQDSFTPVVNSLDPPILLTRVLRTHPSQSWHQIALRMEVLGCEAQDLY 1438
 DB 1367 FONGKVKVQGNQDSFTPVVNSLDPPILLTRVLRTHPSQSWHQIALRMEVLGCEAQDLY 1424

RESULT 13
 AAO18622
 ID AAO18622 standard; Protein; 1424 AA.
 AC AAO18622;
 XX 24-OCT-2002 (first entry)
 DT Human mature B-domainless factor VIII.
 DE Human; factor VIII; fVIII; half-life; mutant; haemophilia;
 KW heparan sulfate proteoglycan-mediated clearance; RAP;
 KW receptor-associated protein; haemostatic; gene therapy;
 KW alpha2 macroglobulin receptor-associated protein.
 XX Homo sapiens.
 OS Key Location/Qualifiers
 FH Domain 373..740
 FT /label= A2 domain
 FT Region 484..509
 FT /label= LRP_binding_region
 XX WO200260951-A2.
 XX 08-AUG-2002.
 XX 11-JAN-2002; 2002WO-US00583.
 XX 12-JAN-2001; 2001US-260904P.
 XX (AMNA-) AMERICAN NAT RED CROSS.
 XX Saenko EL, Sarafanov AG;
 XX WPI; 2002-608501/65.
 XX New mutant factor VIII with reduced sulfate proteoglycan
 PT (HSPG)-dependent or receptor-independent clearance and procoagulant
 PT activity for treating hemophilia -
 XX Claim 9; Fig 12; 161pp; English.
 XX The present invention relates to a mutant factor VIII protein with

CC reduced sulfate proteoglycan (HSPG)-dependent or receptor-independent
 CC clearance and procoagulant activity, which has a nonconservative amino
 CC acid substitution at one or more positions in the A2 domain consisting of
 CC Lys(380, 512, 556, 570 or 659) or Arg(490, 527, 562 or 571) or the C2
 CC domain relative to the wild-type. The mutant factor VIII or the
 CC polynucleotide encoding it and a receptor-associated protein (alpha2
 CC macroglobulin receptor-associated protein or RAP) are useful for treating
 CC haemophilia. The mutated protein has a longer half-life. The present
 CC sequence is the human mature B-domainless factor VIII protein.
 XX

QY Sequence 1424 AA;

Query Match 98.7%; Score 7592; DB 23; Length 1424;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1423; Conservative 0; Mismatches 1; Indels 14; Gaps 1;

QY 1 ATRRYILGAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVYVKTILFVEFTVHLFN 60
 DB 1 ATRRYILGAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVYVKTILFVEFTVHLFN 60
 QY 61 IAKPRPPMGLGPTIOAEVYDVTVTILKNMASHPVSLHAGVSYWKASGAEYDDQTSQ 120
 DB 61 IAKPRPPMGLGPTIOAEVYDVTVTILKNMASHPVSLHAGVSYWKASGAEYDDQTSQ 120
 QY 121 REKEDDKVFPGGSHYVQVLKENGPMASDPLCLITYSLSHVDLVKDLNGLIGALLVCR 180
 DB 121 REKEDDKVFPGGSHYVQVLKENGPMASDPLCLITYSLSHVDLVKDLNGLIGALLVCR 180
 QY 181 EGSIAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVGVYNR 240
 DB 181 EGSIAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVGVYNR 240
 QY 241 SLPLIGCHRSVYVHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQLL 300
 DB 241 SLPLIGCHRSVYVHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQLL 300
 QY 301 MDLQGFLLFCHISHQHDGMEAYVKVDSCEEPQLRMKNBEAEYDDDLTDSMDVYRF 360
 DB 301 MDLQGFLLFCHISHQHDGMEAYVKVDSCEEPQLRMKNBEAEYDDDLTDSMDVYRF 360
 QY 361 DDNSPSFIOIRSVAKKHPTWVHYIAAEEDMDYAPLVLPDDRYSKSOVLNNGPORIG 420
 DB 361 DDNSPSFIOIRSVAKKHPTWVHYIAAEEDMDYAPLVLPDDRYSKSOVLNNGPORIG 420
 QY 421 RYKVKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDILLIIFKQASRPYNIYPHGI 480
 DB 421 RYKVKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDILLIIFKQASRPYNIYPHGI 480
 QY 481 TDVRPLYSRRLPKGVKHLKDPPIIPGEIFKYKWTVTVEDEGTPKSDPCLTRYISGFVME 540
 DB 481 TDVRPLYSRRLPKGVKHLKDPPIIPGEIFKYKWTVTVEDEGTPKSDPCLTRYISGFVME 540
 QY 541 RDLASGLIGPLLCYKESVDQGNQIMSDKRNVLFSVFDENRSMWLTENIQRLFLPAG 600
 DB 541 RDLASGLIGPLLCYKESVDQGNQIMSDKRNVLFSVFDENRSMWLTENIQRLFLPAG 600
 QY 601 VQLEDPEFQASNMHSINGVYFDSLQSLVCLHEVAYWYIISIGAQTDFLSVFFSGYTFKH 660
 DB 601 VQLEDPEFQASNMHSINGVYFDSLQSLVCLHEVAYWYIISIGAQTDFLSVFFSGYTFKH 660
 QY 661 KMVYEDTLTLFPFSGETVFMSENPGWLWILGCHNSDFRNKMTALLKVSSCDKNTGDIYE 720
 DB 661 KMVYEDTLTLFPFSGETVFMSENPGWLWILGCHNSDFRNKMTALLKVSSCDKNTGDIYE 720
 QY 721 DSYEDISAYLLSKNNAIEPRFSQNPVLAKEHOREITRTTILQSDQERIDYDDTISVEMKK 780
 DB 721 DSYEDISAYLLSKNNAIEP-----REITRTTILQSDQERIDYDDTISVEMKK 766
 QY 781 EDFDIYEDENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHVLNRQAQSGSVQFQKVV 840
 DB 767 EDFDIYEDENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHVLNRQAQSGSVQFQKVV 826
 QY 841 FQEFDTGDSFTQPLYRGELNBEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSLSIYEE 900

Db 827 FQFTDGSFTQPLRGELNHLGLLPYIRAEVEDNIMVTFRNOASRPYSFSSLSYBE 886
QY 901 DORQGAEPKPNFKVNETKTYFKVQHHMAPKDEPDCKAWAFSDVDLEKDVHSLGIP 960
Db 887 DORQGAEPKPNFKVNETKTYFKVQHHMAPKDEPDCKAWAFSDVDLEKDVHSLGIP 946
QY 961 LLVCHTNTLNPAHGRQVTVQEPALFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTPK 1020
Db 947 LLVCHTNTLNPAHGRQVTVQEPALFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTPK 1006
QY 1021 NYRFAHNGIYIMDTLPCILYMAQDORIRWYLLSGNSNENIHSIFSGHVTVRKBEYKMA 1080
Db 1007 NYRFAHNGIYIMDTLPCILYMAQDORIRWYLLSGNSNENIHSIFSGHVTVRKBEYKMA 1066
QY 1081 LYNLYPGVETVEMLPKAGIWRVECLIGEHLAGMSTLFLVYSNKCQTPGLWASGHID 1140
Db 1067 LYNLYPGVETVEMLPKAGIWRVECLIGEHLAGMSTLFLVYSNKCQTPGLWASGHID 1126
QY 1141 FOITASQYQGWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGARQKFS 1200
Db 1127 FOITASQYQGWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGARQKFS 1186
QY 1201 SYIISQFIIMYSLDGKKWQTYRGNSTGTLWVFGNVDSGKHNIFNPPIIARYIRLHPT 1260
Db 1187 SYIISQFIIMYSLDGKKWQTYRGNSTGTLWVFGNVDSGKHNIFNPPIIARYIRLHPT 1246
QY 1261 HYSIRSTLWELMGCNLNSCPLMGESKAISDAQITASSYFTNMFATWSPSKARLHLQ 1320
Db 1247 HYSIRSTLWELMGCNLNSCPLMGESKAISDAQITASSYFTNMFATWSPSKARLHLQ 1306
QY 1321 RSNARPOVNNPKEMLVQDFQKTMKVTGVTQGVKSLLTSMYVKEFLISSQDGHQWTLF 1380
Db 1307 RSNARPOVNNPKEMLVQDFQKTMKVTGVTQGVKSLLTSMYVKEFLISSQDGHQWTLF 1366
QY 1381 FQNGKVKVQGNQDSFTPVVNSLDPLLRILYLRHFPQSWHQAIRMEVLGCEAQDLY 1438
Db 1367 FQNGKVKVQGNQDSFTPVVNSLDPLLRILYLRHFPQSWHQAIRMEVLGCEAQDLY 1424

RESULT 14
ABG92541
ID ABG92541 standard; Protein; 1447 AA.
XX AC ABG92541;
XX DT 19-NOV-2002 (first entry)
XX DE 5Arg B-domain-deleted-Factor VIII (FVIII).
XX KW Human; alpha-galactosidase; Factor VIII; Factor IX;
XX KW gene therapy.
XX OS Homo sapiens.
XX OS Synthetic.
XX PN WO200264799-A2.
XX PD 22-AUG-2002.
XX PF 11-OCT-2001; 2001WO-US42655.
XX PR 11-OCT-2000; 2000US-0686497.
XX XX (TRAN-) TRANSKARYOTIC THERAPIES INC.
XX XX Seldon RF, Miller AM, Treco DS;
XX XX WPI; 2002-627600/67.
XX XX N-PSDB; ABS68000.
XX XX New synthetic nucleic acid sequence involving a continuous stretch of a
PT least 150 common codons, useful for expressing mammalian, preferably

PT human proteins e.g. alpha-galactosidase or Factor VIII or IX or for
PT gene therapy -
XX
PS Example 1; Figure 9; 115pp; English.
XX
CC The invention relates to a synthetic nucleic acid sequence, where at
CC least one non-common or less-common codon is replaced with a common
CC codon. The synthetic nucleic acid is useful for expressing mammalian,
CC preferably human proteins e.g. alpha-galactosidase or Factor VIII or IX
CC or for gene therapy. The synthetic nucleic acid allows precise dosing and
CC reduces treatment costs. It is simple to apply in treating patients and
CC is curative (one gene therapy treatment has the potential to last a
CC patient's lifetime. ABG92540-ABG92541 represent Factor VIII and Factor
XX IX synthetic amino acid sequences of the invention.
SQ Sequence 1447 AA;
Query Match 98.7%; Score 7588; DB 23; Length 1447;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1423; Conservative 1; Mismatches 4; Indels 10; Gaps 1;
QY 1 ATRRYLGAVELSWDMQSDLGELPVDARFPFPRPKSPFPNTSVVYKKTLPVEFTVHLFN 60
Db 20 ATRRYLGAVELSWDMQSDLGELPVDARFPFPRPKSPFPNTSVVYKKTLPVEFTVHLFN 79
QY 61 IAKPRPPNMGLLGPTIOAEVYDVTVTILKNMASHPVSLHAGVSVYKASEGAEYDDQTSQ 120
Db 80 IAKPRPPNMGLLGPTIOAEVYDVTVTILKNMASHPVSLHAGVSVYKASEGAEYDDQTSQ 139
QY 121 REKEDKVPFGSGSHYVQVLKENGPMASDPLCLTYSYLSHVLDLKNLSGLIGALLVCR 180
Db 140 REKEDKVPFGSGSHYVQVLKENGPMASDPLCLTYSYLSHVLDLKNLSGLIGALLVCR 199
QY 181 EGS�AKEKTQTLHKILLFAVDEGKSWHSETKNSLMQDRDAASARAPQMTWVYVNR 240
Db 200 EGS�AKEKTQTLHKILLFAVDEGKSWHSETKNSLMQDRDAASARAPQMTWVYVNR 259
QY 241 SIPLGLIGCHRSVYWHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLL 300
Db 260 SIPLGLIGCHRSVYWHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLL 319
QY 301 MDLQGLFLFCHISSHQHDGMEAYVVDSCPEPQLRMKNNNEAEYDDDLTSEMDVVRF 360
Db 320 MDLQGLFLFCHISSHQHDGMEAYVVDSCPEPQLRMKNNNEAEYDDDLTSEMDVVRF 379
QY 361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEDWDYAPLAPDDSYKSOYLNNGPORIG 420
Db 380 DDNSPSFIQIRSVAKKPKTWVHYIAAEEDWDYAPLAPDDSYKSOYLNNGPORIG 439
QY 421 RYKVKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHGI 480
Db 440 RYKVKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHGI 499
QY 481 TDVRLYLRRLPKGVKHLKDFPILGEPFKYKWTVTVEDGPTKSDPRCLTRYSSPVNME 540
Db 500 TDVRLYLRRLPKGVKHLKDFPILGEPFKYKWTVTVEDGPTKSDPRCLTRYSSPVNME 559
QY 541 RDLASGLIGPLLICVKESVDQRGNQIMSDKRNVLFSVFDENKSWLTENIORFLPNPAG 600
Db 560 RDLASGLIGPLLICVKESVDQRGNQIMSDKRNVLFSVFDENKSWLTENIORFLPNPAG 619
QY 601 VOLEDPEFOASNIHMSINGVYFDSQLSVCLHEVAYWYILSIAQTDPLFSVFGVTEKH 660
Db 620 VOLEDPEFOASNIHMSINGVYFDSQLSVCLHEVAYWYILSIAQTDPLFSVFGVTEKH 679
QY 661 KMYEDTTLFPFSGETVPMSENPGMLILGCHNSDFRNRGMTALLKVSSCDKNTGDYXE 720
Db 680 KMYEDTTLFPFSGETVPMSENPGMLILGCHNSDFRNRGMTALLKVSSCDKNTGDYXE 739
QY 721 DSYVEDISAYLLSKNNAIEPRSPQPPVLKQHREITRTLQSDQEEIDYDDTISVEMKK 780
Db 740 DSYVEDISAYLLSKNNAIEPRSPQPPVLKQHREITRTLQSDQEEIDYDDTISVEMKK 789

Qy 781 EDFDIYDEDENSPRSQKTRHYFIAAVERLMDYGMSSSPHVLNRNQAQSGVDFQKVV 840
Db 790 EDFDIYDEDENSPRSQKTRHYFIAAVERLMDYGMSSSPHVLNRNQAQSGVDFQKVV 849
Qy 841 FQETDGSFTQPLRGELNEHLGLGYPYIRAEVEDNIMVFRNQAQSPYFYSSLISYE 900
Db 850 FQETDGSFTQPLRGELNEHLGLGYPYIRAEVEDNIMVFRNQAQSPYFYSSLISYE 909
Qy 901 DORQGAERKMFVKNPTKTVFWKQVHMAPTKDEDFCKAWAYFSDVDLEKDVHSLIGP 960
Db 910 DORQGAERKMFVKNPTKTVFWKQVHMAPTKDEDFCKAWAYFSDVDLEKDVHSLIGP 969
Qy 961 LLVCHTNTLNAHQGRQVTVQSFALFFTFIDETKSWYFTENNERCAPNQMEDPTFKE 1020
Db 970 LLVCHTNTLNAHQGRQVTVQSFALFFTFIDETKSWYFTENNERCAPNQMEDPTFKE 1029
Qy 1021 NYRFAHNGYIMDTLPGLVMAQDQRIRWYLLSMGSENENIHSIHFSGHVTVRKKEBYMA 1080
Db 1030 NYRFAHNGYIMDTLPGLVMAQDQRIRWYLLSMGSENENIHSIHFSGHVTVRKKEBYMA 1089
Qy 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYVSNKQOTPLGMASGHIRD 1140
Db 1090 LYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYVSNKQOTPLGMASGHIRD 1149
Qy 1141 FQITASGOYQGWAPKFLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGINKTGARQKFS 1200
Db 1150 FQITASGOYQGWAPKFLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGINKTGARQKFS 1209
Qy 1201 SLYISQFIIMYSLGKKWQTVRGNSTGTLMVFPNGVDSSGKKNIFNPPITARIYRLHPT 1260
Db 1210 SLYISQFIIMYSLGKKWQTVRGNSTGTLMVFPNGVDSSGKKNIFNPPITARIYRLHPT 1269
Qy 1261 HYSIRSTLWELMGCGLNCSMPGMSKALSDAQITASSYFTNWEATWSPSKARLHLQG 1320
Db 1270 HYSIRSTLWELMGCGLNCSMPGMSKALSDAQITASSYFTNWEATWSPSKARLHLQG 1329
Qy 1321 RSNARWQVNNPKEWLQVDFQTKMKTGVTGVTQGVKSLTSMYVKEFLISSQDGHQWTLF 1380
Db 1330 RSNARWQVNNPKEWLQVDFQTKMKTGVTGVTQGVKSLTSMYVKEFLISSQDGHQWTLF 1389
Qy 1381 PQNGKVKVQGNQDSFTPVNSLDPPLLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY 1438
Db 1390 PQNGKVKVQGNQDSFTPVNSLDPPLLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY 1447

RESULT 15
AAP80267
ID AAP80267 standard; protein; 1425 AA.
XX
AC AAP80267;
XX
DT 25-MAR-2003 (updated)
DT 10-OCT-1990 (first entry)
XX
DE Modified factor VIII:C sequence with the R740-E1649 deletion.
DE
KW Modified factor VIII:C; maturation polypeptide; haemophilia;
KW blood coagulation; RE deletion.
XX
OS Homo sapiens.
XX
EN W08800831-A.
XX
PD 11-FEB-1988.
XX
PF 31-JUL-1987; 87WO-US01814.
XX
PR 01-AUG-1986; 86US-0893375.
XX
PA (BIOJ) BIOGEN NV.
PA (PASE/) PASEK M P.
XX
PI Pasek MP;

XX WPI; 1988-049866/07.
DR N-PSDB; AAN80446.
XX
PT New DNA sequences encoding modified factor VIII:C - with deletion of DNA
XX encoding maturation polypeptide, useful for high yield transformation.
PS Claim 3; Page 57-58-59-60; 97pp; English.
XX
CC The entire sequence encoding the maturation polypeptide of
CC factor VIII:C is deleted, i.e. Arg 740-Glu 1649.
CC The full length factor VIII:C cDNA has two changes with respect to the
CC published sequence (EPO application 160457):
CC CTG to CTA at Leu 242 and TTC to CTC change at amino acid residue 1880
CC (Phe to Leu). The product is produced in approx. 20 times higher
CC yields than previous recombinant produced factor VIII:C and are more
CC easily purified. The peptide is used for treating haemophilia A, both
CC acute and prolonged bleeding.
CC See also AAN80444 and AAN80447.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 1425 AA;
Query Match 98.6%; Score 7581; DB 9; Length 1425;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1421; Conservative 0; Mismatches 3; Indels 14; Gaps 1;
Qy 1 ATRRYYLGAVELSDYMQSDLGELPVDARPPRPVPKSPFNTSVVYKTLFVRETVHLEN 60
Db 2 ATRRYYLGAVELSDYMQSDLGELPVDARPPRPVPKSPFNTSVVYKTLFVRETVHLEN 61
Qy 61 IAKPRPPWMLGLPTIAQAEVYDVTITLKNMASHPVSLHAGVSYMKASGAEYDDQTSQ 120
Db 62 IAKPRPPWMLGLPTIAQAEVYDVTITLKNMASHPVSLHAGVSYMKASGAEYDDQTSQ 121
Qy 121 REKEDDKVFPQSGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
Db 122 REKEDDKVFPQSGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 181
Qy 181 EGSILAKETOTLHKFILLFAVDFGKSWHSETKNSLMODRDAASARAWPKMHTVGVYNR 240
Db 182 EGSILAKETOTLHKFILLFAVDFGKSWHSETKNSLMODRDAASARAWPKMHTVGVYNR 241
Qy 241 SLPLGLIGCHRSVYVHWVIGMGTTPFVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL 300
Db 242 SLPLGLIGCHRSVYVHWVIGMGTTPFVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL 301
Qy 301 MDLQGFLLFCHISSHQHDMGMEAYVYKVDSCPEEPQRLMKNNNEEAEDYDDDLTDSEMDVVR 360
Db 302 MDLQGFLLFCHISSHQHDMGMEAYVYKVDSCPEEPQRLMKNNNEEAEDYDDDLTDSEMDVVR 361
Qy 361 DDNSPSPIQIRSVAKKHPTKTHVYIAAEEDWDYAPLVLPADDRSYKSOVLNNGPORG 420
Db 362 DDNSPSPIQIRSVAKKHPTKTHVYIAAEEDWDYAPLVLPADDRSYKSOVLNNGPORG 421
Qy 421 RYKVKVFMAYTDETFKTRAIQHSIGLPLGELVGEVGDITLLIIFKNQASPNYNYPHGI 480
Db 422 RYKVKVFMAYTDETFKTRAIQHSIGLPLGELVGEVGDITLLIIFKNQASPNYNYPHGI 481
Qy 481 TDVRLYSRRLPKGVKHLKDFPILPGBIFKYKWTVTVEDGPKSDPCLTRYSYFVNM 540
Db 482 TDVRLYSRRLPKGVKHLKDFPILPGBIFKYKWTVTVEDGPKSDPCLTRYSYFVNM 541
Qy 541 RDLASGLIGPLLCYKESVDORGQINMSDKRNVILFVFDENRSWYLTENIQRLPNPAG 600
Db 542 RDLASGLIGPLLCYKESVDORGQINMSDKRNVILFVFDENRSWYLTENIQRLPNPAG 601
Qy 601 VQLEDPEFQASNMHSINGVYVDFSLQSVCLHEVAYVYILSIGAQDTFLSFYFSGYTFKH 660
Db 602 VQLEDPEFQASNMHSINGVYVDFSLQSVCLHEVAYVYILSIGAQDTFLSFYFSGYTFKH 661
Qy 661 KMVYEDTLTLPFPFSGETVFMSEMPGLWILGCHNSDFNRNGTALLKVSSCDKNTGDYFE 720

Db 662 KMVYEDTTLPPFSGETVFMSENFGMLWILGCHNSDFRNRGNTALLKVSSCDKNTGDIYE 721
QY 721 DSYEDISAYLLSKNNAIEPRSPQPPVLKRHRQREITRTTQSDQOEEDIDYDITISVEMKK 780
Db 722 DSYEDISAYLLSKNNAIEP-----REITRTTQSDQOEEDIDYDITISVEMKK 767
QY 781 EDFDIYDEENQSPRSFQKTRHYFIAAVERLWDYGMSSSHVLRNRAQSGSVPOFKKV 840
Db 768 EDFDIYDEENQSPRSFQKTRHYFIAAVERLWDYGMSSSHVLRNRAQSGSVPOFKKV 827
QY 841 FOEFTDGSFTQPLYGELNEHGLAGPYIRAEVEDNIMVTFRNOASRPVSYSSLSIYEE 900
Db 828 FOEFTDGSFTQPLYGELNEHGLAGPYIRAEVEDNIMVTFRNOASRPVSYSSLSIYEE 887
QY 901 DORQGAEPKRFVNEKNTKYFWKHMAPTKDFCKAWAYFSDVDLEKDVHSLGIP 960
Db 888 DORQGAEPKRFVNEKNTKYFWKHMAPTKDFCKAWAYFSDVDLEKDVHSLGIP 947
QY 961 LLVCHTNTLNPAGHQVTVQEPALFTTIDETKSWYFTENMERNCRAPCNIQMEDPTTKE 1020
Db 948 LLVCHTNTLNPAGHQVTVQEPALFTTIDETKSWYFTENMERNCRAPCNIQMEDPTTKE 1007
QY 1021 NYRFAINGYIMDTLPGLVMAODQIRWYLLSGNSNENIHSIFSGHVFTVRKKEYYKMA 1080
Db 1008 NYRFAINGYIMDTLPGLVMAODQIRWYLLSGNSNENIHSIFSGHVFTVRKKEYYKMA 1067
QY 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPPLGMAASHIRD 1140
Db 1068 LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPPLGMAASHIRD 1127
QY 1141 FOITASQYQGWAPKLARLHYSGSINAWSTKEPPFSWKVDLLAPMIIHGKIQGAROKFS 1200
Db 1128 FOITASQYQGWAPKLARLHYSGSINAWSTKEPPFSWKVDLLAPMIIHGKIQGAROKFS 1187
QY 1201 SLYISQFIIMYSLDGKKWOTYRGNSTGTLWVFFGNVDSSGIGKHNIFNPPIIARYIRLHPT 1260
Db 1188 SLYISQFIIMYSLDGKKWOTYRGNSTGTLWVFFGNVDSSGIGKHNIFNPPIIARYIRLHPT 1247
QY 1261 HYSIRSTLRMELMGCDLNSCMLPGMESKAI SDAQITASSYFTNNPATWSPSKARHLQ 1320
Db 1248 HYSIRSTLRMELMGCDLNSCMLPGMESKAI SDAQITASSYFTNNPATWSPSKARHLQ 1307
QY 1321 RSNARWPQVNNPKEMLQVDFQKTMKVTGTTQGVKSLLTSMYVKEFLISSSQDGHQWTLF 1380
Db 1308 RSNARWPQVNNPKEMLQVDFQKTMKVTGTTQGVKSLLTSMYVKEFLISSSQDGHQWTLF 1367
QY 1381 FONGKVKVFGNQDSFTPVVNSLDPLLRILRIHPQSWHQAIALRMEVLGCEAODLY 1438
Db 1368 FONGKVKVFGNQDSFTPVVNSLDPLLRILRIHPQSWHQAIALRMEVLGCEAODLY 1425

Search completed: December 9, 2003, 16:53:08
Job time : 67 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 16:52:06 ; Search time 22 Seconds
(without alignments)
2765.592 Million cell updates/sec

Title: US-10-006-091-1

Perfect score: 7691

Sequence: 1 ATRRYLGAVELSWDMQSD.....VWQIALRMEVLGCEAQLY 1438

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A.COMB.pap.*
- 2: /cgn2_6/ptodata/1/iaa/5B.COMB.pap.*
- 3: /cgn2_6/ptodata/1/iaa/6A.COMB.pap.*
- 4: /cgn2_6/ptodata/1/iaa/6B.COMB.pap.*
- 5: /cgn2_6/ptodata/1/iaa/PTUS.COMB.pap.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7691	100.0	1438	4	US-09-209-916-1
2	7674	99.8	1471	1	US-08-683-839B-3
3	7434	96.7	1661	2	US-08-882-083-2
4	7434	96.7	1661	2	US-08-558-107-2
5	7434	96.7	1661	3	US-09-243-539-2
6	7234	94.1	2332	1	US-07-864-004B-4
7	7234	94.1	2332	1	US-08-251-937A-4
8	7234	94.1	2332	1	US-08-212-133A-2
9	7234	94.1	2332	1	US-08-474-503-2
10	7234	94.1	2332	1	US-08-670-707A-2
11	7234	94.1	2332	3	US-09-037-601-2
12	7234	94.1	2332	4	US-09-315-179-2
13	7234	94.1	2332	4	US-09-523-656-2
14	7234	94.1	2332	5	PCT-US93-03275-4
15	7234	94.1	2332	5	PCT-US94-13200-2
16	7234	94.1	2351	1	US-08-121-202-2
17	7227	94.0	2351	1	US-08-366-851A-2
18	7225	93.9	2351	6	542260-1
19	7224	93.9	2332	1	US-08-276-594A-2
20	7221	93.9	2351	6	5171844-2
21	7195	93.6	2332	3	US-09-324-867-3
22	6560	85.3	1467	4	US-09-523-656-38
23	6489	84.4	1443	2	US-08-670-707A-39
24	6489	84.4	1443	3	US-09-037-601-39
25	6489	84.4	1443	4	US-09-315-179-39
26	6272	81.5	2343	3	US-09-324-867-2
27	6227	81.0	2133	2	US-08-670-707A-37

ALIGNMENTS

RESULT 1

US-09-209-916-1
; Sequence 1, Application US/09209916
; Patent No. 6358703
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255
; CURRENT APPLICATION NUMBER: US/09/209,916
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-09-209-916-1

Query Match	100.0%;	Score	7691;	DB	4;	Length	1438;
Best Local Similarity	100.0%;	Pred. No.	0;	Mismatches	0;	Indels	0;
Matches	1438;	Conservative	0;				
Qy	1	ATRRYLGAVELSWDMQSDLGELPVDARPPRPVKSPFFNTSVVYKKTFLVFEFTVHLFN	60				
Db	1	ATRRYLGAVELSWDMQSDLGELPVDARPPRPVKSPFFNTSVVYKKTFLVFEFTVHLFN	60				
Qy	61	IAKPRPPMGLGPTIQAEVYDVTVTILKNMASHPVSLHVGVSYWKASGEAYDDQTSQ	120				
Db	61	IAKPRPPMGLGPTIQAEVYDVTVTILKNMASHPVSLHVGVSYWKASGEAYDDQTSQ	120				
Qy	121	REKEDDKVFPGSGSHYVQVVKENGPMASDPLCLTYSYLSHVLDVKDLSGLIGALLVCR	180				
Db	121	REKEDDKVFPGSGSHYVQVVKENGPMASDPLCLTYSYLSHVLDVKDLSGLIGALLVCR	180				
Qy	181	EGSLAKETQTLHKFILLFAVFDGKSWHSSTKNSLMQDRDAASARAMPKMTVNGVYNR	240				
Db	181	EGSLAKETQTLHKFILLFAVFDGKSWHSSTKNSLMQDRDAASARAMPKMTVNGVYNR	240				
Qy	241	SLPGLIGCHRSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL	300				
Db	241	SLPGLIGCHRSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL	300				
Qy	301	MDLGQFLFCHSHSHQHDGMAYVKVDSCPREPQLRMKNNEAEYDDDLTDSEMDVYRF	360				
Db	301	MDLGQFLFCHSHSHQHDGMAYVKVDSCPREPQLRMKNNEAEYDDDLTDSEMDVYRF	360				

Db 301 MDLQGFLLFCHISSHQHDGMEAYVKVDSCEPFPQLRMKNNEEAEDYDDDLTDSEMDVVRP 360
QY 361 DDNSPSFIQIRSAKPKTKWHVYIAAEEEDWDAPLVADPDRSYKSOYLNNGPQRIG 420
Db 361 DDNSPSFIQIRSAKPKTKWHVYIAAEEEDWDAPLVADPDRSYKSOYLNNGPQRIG 420
QY 421 RYKVKRPMAYTDETFKTRAIQHESGILGPLLYGEVGTLLIIFKNQASRPYNTYPHGI 480
Db 421 RYKVKRPMAYTDETFKTRAIQHESGILGPLLYGEVGTLLIIFKNQASRPYNTYPHGI 480
QY 481 TVRPLYSRRLPKGVKHLKDFILPGEIFKYKWTVTVEDGPKSDRCLTRYSSFFVNM 540
Db 481 TVRPLYSRRLPKGVKHLKDFILPGEIFKYKWTVTVEDGPKSDRCLTRYSSFFVNM 540
QY 541 RDLASGLIPLLI CYKESVDQGNQIMSDKRNVLIFSVDENRSWLTENIQRFNPAG 600
Db 541 RDLASGLIPLLI CYKESVDQGNQIMSDKRNVLIFSVDENRSWLTENIQRFNPAG 600
QY 601 VQLEDPEFOASIMHSINGVYFDSLSQVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660
Db 601 VQLEDPEFOASIMHSINGVYFDSLSQVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660
QY 661 KMVEDTLTLFPFSGTVMFMSNPGMLWILGCHNSDFRNGMTALLKVSCKDXTGDIYE 720
Db 661 KMVEDTLTLFPFSGTVMFMSNPGMLWILGCHNSDFRNGMTALLKVSCKDXTGDIYE 720
QY 721 DSYEDISAVLLSKNNAIEPRSPONPPVLKRQREITRTTLOSQREIDYDDTISVEMKK 780
Db 721 DSYEDISAVLLSKNNAIEPRSPONPPVLKRQREITRTTLOSQREIDYDDTISVEMKK 780
QY 781 EPDFIYDEENOSPRSFQKTRHYFIAAVERLWDYGMSSSPHYLRNRAQSGVPOPKVY 840
Db 781 EPDFIYDEENOSPRSFQKTRHYFIAAVERLWDYGMSSSPHYLRNRAQSGVPOPKVY 840
QY 841 FQFTDGSFTQPIYRGELNEHLGLGPYIRAEVEDNIMVTFRNOASRPYSFSSLSIYEE 900
Db 841 FQFTDGSFTQPIYRGELNEHLGLGPYIRAEVEDNIMVTFRNOASRPYSFSSLSIYEE 900
QY 901 DQOQGAEPKRFVKNPNETKYFWKVQHMAPTKDEFCKAWAYFSDVLEKDVHSLGIP 960
Db 901 DQOQGAEPKRFVKNPNETKYFWKVQHMAPTKDEFCKAWAYFSDVLEKDVHSLGIP 960
QY 961 LLVCHTNTLNPAGROVTVQEFALFTTIFDETSKWTFTENMERNCRAPCNIQMEDPTFK 1020
Db 961 LLVCHTNTLNPAGROVTVQEFALFTTIFDETSKWTFTENMERNCRAPCNIQMEDPTFK 1020
QY 1021 NYRFAHNGVIMDTLGLVMAQOQIRWYLLSGNSNENIHSIFHSGHVTVRKKEKYKA 1080
Db 1021 NYRFAHNGVIMDTLGLVMAQOQIRWYLLSGNSNENIHSIFHSGHVTVRKKEKYKA 1080
QY 1081 LYNLYPGVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCOTPLGNASGHIRD 1140
Db 1081 LYNLYPGVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCOTPLGNASGHIRD 1140
QY 1141 FQITASQYQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHIGIKTOGARQFS 1200
Db 1141 FQITASQYQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHIGIKTOGARQFS 1200
QY 1201 SLVISOPIIMYSLDGKKWQYTRNGSTGLTVFFGNVDSSGINKNINPPIIARYIRLHPT 1260
Db 1201 SLVISOPIIMYSLDGKKWQYTRNGSTGLTVFFGNVDSSGINKNINPPIIARYIRLHPT 1260
QY 1261 HYSIRSTRMELMCDLNSCMPLGMSKAI SDAQITASSVFTNMFATWSPSKARLHLOQ 1320
Db 1261 HYSIRSTRMELMCDLNSCMPLGMSKAI SDAQITASSVFTNMFATWSPSKARLHLOQ 1320
QY 1321 RSNARWQVANNPKLEQVDFQKTMKVTVGTITQGVKSLTSMYKVEFLISSQDGHQWTLF 1380
Db 1321 RSNARWQVANNPKLEQVDFQKTMKVTVGTITQGVKSLTSMYKVEFLISSQDGHQWTLF 1380
QY 1381 FQNGKVKVFGNQDSFTFVNVNSLDPPLLTRYLRHHPQSVWHQIALRMEVLGCAQDILY 1438
Db 1381 FQNGKVKVFGNQDSFTFVNVNSLDPPLLTRYLRHHPQSVWHQIALRMEVLGCAQDILY 1438

RESULT 2

US-08-839B-3
; Sequence 3, Application US/08683839B
; Patent No. 574326

GENERAL INFORMATION:

APPLICANT: Tll, Charles . R. et al.
TITLE OF INVENTION: Use of Viral Cis-Acting Post-Transcriptional
TITLE OF INVENTION: Regulatory Sequences To Increase Expression of
TITLE OF INVENTION: Introns Genes Containing Near-Consensus Splice Sites
NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

APPLICANT DATA:

APPLYING DATE: 11-MARCH-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/683.839B

FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Remillard, Jane E.
REGISTRATION NUMBER: 38,872
REFERENCE/DOCKET NUMBER: TTI-138
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 1471 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-683-839B-3

Query Match 99.8%; Score 7674; DB 1; Length 1471;

Best Local Similarity 99.0%; Pred. No. 0;

Matches 1438; Conservative 0; Mismatches 0; Indels 14; Gaps 1;

QY 1 ATRRYLGAVELSWDYMQSDLGELPVDARPPRPVKSPFPNTSVVYKKTLFVEFTVHLFN 60
Db 20 ATRRYLGAVELSWDYMQSDLGELPVDARPPRPVKSPFPNTSVVYKKTLFVEFTVHLFN 79
QY 61 IAKPRPPMGLLGPTIOAEVYTVITLQKMAHPVSLHAGVSVTWKASEGAEYDDQTSQ 120
Db 80 IAKPRPPMGLLGPTIOAEVYTVITLQKMAHPVSLHAGVSVTWKASEGAEYDDQTSQ 139
QY 121 RKEEDDKVPFGSGHTYVWQVLKENGPMASDPLCTLYSLSHVDLVKDLNSGLIGALLVCR 180
Db 140 RKEEDDKVPFGSGHTYVWQVLKENGPMASDPLCTLYSLSHVDLVKDLNSGLIGALLVCR 199
QY 181 EGSLAKEKQTQLHKFTLLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
Db 200 EGSLAKEKQTQLHKFTLLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 259
QY 241 SLPLGLGCHRKSVYHVIQMGTTPEVHSIFLEGHTFLVNRHQASLEISPIFLTAQTLL 300
Db 260 SLPLGLGCHRKSVYHVIQMGTTPEVHSIFLEGHTFLVNRHQASLEISPIFLTAQTLL 319
QY 301 MDLQGFLLFCHISSHQHDGMEAYVKVDSCEPFPQLRMKNNEEAEDYDDDLTDSEMDVVRP 360
Db 320 MDLQGFLLFCHISSHQHDGMEAYVKVDSCEPFPQLRMKNNEEAEDYDDDLTDSEMDVVRP 379

QY 361 DDNSPSFIQIRSVAKKPKTKTWYHIAABEDMDYAPLVLPADDRSKSYQLNNGPORIG 420
DB 380 DDNSPSFIQIRSVAKKPKTKTWYHIAABEDMDYAPLVLPADDRSKSYQLNNGPORIG 439
QY 421 RYKVKVFMAYTDETFKTRAIQESGILGPLLGEVGDITLLIFKNQASRPYNIYPHGI 480
DB 440 RYKVKVFMAYTDETFKTRAIQESGILGPLLGEVGDITLLIFKNQASRPYNIYPHGI 499
QY 481 TDVRPLYSRRLPKGVKHLKDPFLLPGBI PKYKWTVTVEDGPTKSDPCLTRYSSFYNME 540
DB 500 TDVRPLYSRRLPKGVKHLKDPFLLPGBI PKYKWTVTVEDGPTKSDPCLTRYSSFYNME 559
QY 541 RDLASGLIGPLLICYKESVDORGNOIMSDKKNVILFSVFDENRSWYLTENIQRLPNPAG 600
DB 560 RDLASGLIGPLLICYKESVDORGNOIMSDKKNVILFSVFDENRSWYLTENIQRLPNPAG 619
QY 601 VQLEDPFQASNMHSNGVYVDFSLQLSVCLHEVAYWYILSIGACTDPLSVFFSGYTFKH 660
DB 620 VQLEDPFQASNMHSNGVYVDFSLQLSVCLHEVAYWYILSIGACTDPLSVFFSGYTFKH 679
QY 661 KQVYEDTILTFPPSGEIVFMSMENPGLMILGCHNSDFRNRGMTALLKVSSCDKNTGDIYE 720
DB 680 KQVYEDTILTFPPSGEIVFMSMENPGLMILGCHNSDFRNRGMTALLKVSSCDKNTGDIYE 739
QY 721 DSYEDI SAYLLSKNNAIEPRFSQN-----PPVLKXHQREITRTTLOSQOE 766
DB 740 DSYEDI SAYLLSKNNAIEPRFSQNRSRHPSTROKFNATPPVLKXHQREITRTTLOSQOE 799
QY 767 EIDYDDTISVEMKEDFDIYDEDNOSPRSKTRHYFIAAVERLMDYGMSSPHVLRN 826
DB 800 EIDYDDTISVEMKEDFDIYDEDNOSPRSKTRHYFIAAVERLMDYGMSSPHVLRN 859
QY 827 RAQSGSVPOFKVYFQEFDSGFTQPLVRGELNEHLGLLGPYIRAEVEDNIMVTFRNQAS 886
DB 860 RAQSGSVPOFKVYFQEFDSGFTQPLVRGELNEHLGLLGPYIRAEVEDNIMVTFRNQAS 919
QY 887 RPYSFYSSLSIYEBDQOGAPRKNFVKPNETKTYFMKVQHMAPTKDEPCKAWAYPSD 946
DB 920 RPYSFYSSLSIYEBDQOGAPRKNFVKPNETKTYFMKVQHMAPTKDEPCKAWAYPSD 979
QY 947 VLEKDVHSGILGPLLCHTNTLANPAGROVTVQEPALFFTFDETKSWYFTEWNERCR 1006
DB 980 VLEKDVHSGILGPLLCHTNTLANPAGROVTVQEPALFFTFDETKSWYFTEWNERCR 1039
QY 1007 APCNIQMEDPTFKENYRPHAINGYIMDTLPLGVMAQDQIRWYLLSMGNSNENIHSIFSG 1066
DB 1040 APCNIQMEDPTFKENYRPHAINGYIMDTLPLGVMAQDQIRWYLLSMGNSNENIHSIFSG 1099
QY 1067 HVFTVRKKEEYKMALYNLYPGVFFTEVEMLPKAGIRVRVECLIGEHLHAGMSTLFLVYSNK 1126
DB 1100 HVFTVRKKEEYKMALYNLYPGVFFTEVEMLPKAGIRVRVECLIGEHLHAGMSTLFLVYSNK 1159
QY 1127 CQPLGMASGHIRDFQITASGOYGQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMI 1186
DB 1160 CQPLGMASGHIRDFQITASGOYGQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMI 1219
QY 1187 IHGIKTQAROKFSSLYISQFIIMYSLDGKKWQYRGNSTGTLMVFFGNVDSGGIKNIF 1246
DB 1220 IHGIKTQAROKFSSLYISQFIIMYSLDGKKWQYRGNSTGTLMVFFGNVDSGGIKNIF 1279
QY 1247 NPPIIARIYRLHPTHTYSIRSTRMELMGCDLNSCMPLGMSKASDAQITASSYFTNMF 1306
DB 1280 NPPIIARIYRLHPTHTYSIRSTRMELMGCDLNSCMPLGMSKASDAQITASSYFTNMF 1339
QY 1307 ATWSPSKARLHQGRSNARPOVNNPKWLQVDFQKTMKVTVQVTKVSKLLTSMYKKEF 1366
DB 1340 ATWSPSKARLHQGRSNARPOVNNPKWLQVDFQKTMKVTVQVTKVSKLLTSMYKKEF 1399
QY 1367 LISSODGHWMTLFFQNGKVKYFQGNQDSFTPVNSLDPPLLTRYLRIHPQSWHQAIALR 1426
DB 1400 LISSODGHWMTLFFQNGKVKYFQGNQDSFTPVNSLDPPLLTRYLRIHPQSWHQAIALR 1459

QY 1427 MEVLGCEAODLY 1438
DB 1460 MEVLGCEAODLY 1471
RESULT 3
US-08-882-083-2
; Sequence 2, Application US/08882083
; Patent No. 5863292
; GENERAL INFORMATION:
; APPLICANT: VORBERG, Johannes J.
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,083
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/558,107
; FILING DATE: 13-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ISACSON, John P.
; REGISTRATION NUMBER: 33,715
; REFERENCE/DOCKET NUMBER: 30472/212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1661 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-882-083-2

Query Match 96.7%; Score 7434; DB 2; Length 1661;
Best Local Similarity 86.0%; Pred. No. 0;
Matches 1418; Conservative 4; Mismatches 10; Indels 216; Gaps 4;

QY 1 ATRRYVLGAVLSWDMYQSDIGELPVDARFPVRPKSPFNTSVVYKKTFLVEFTVHLFN 60
DB 20 ATRRYVLGAVLSWDMYQSDIGELPVDARFPVRPKSPFNTSVVYKKTFLVEFTVHLFN 79
QY 61 IAKPRPPMGLLGPTIOAEVYDVTVTILKNASHPVSILHVGVSYWKASGAEYDDQTSQ 120
DB 80 IAKPRPPMGLLGPTIOAEVYDVTVTILKNASHPVSILHVGVSYWKASGAEYDDQTSQ 139
QY 121 REKEDDKVFPQSGSYWYQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
DB 140 REKEDDKVFPQSGSYWYQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199
QY 181 EGS�AKEKTOTLHKFILLFAVDFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
DB 200 EGS�AKEKTOTLHKFILLFAVDFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 259
QY 241 SLPLGIGCHRSKVYWHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPTTFLTAQTLL 300
DB 260 SLPLGIGCHRSKVYWHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPTTFLTAQTLL 319
QY 301 MDLQOFLFLFCHISSHQHDGMEAYVKVDSCPREPQLRMKNNEAEYDDDLTDSMDVVRP 360

Db 320 MDLGGFLFCHISSHQHDGMEAYVVDSCPEBPQLRMKNEBEAEDDDDLTSEMDVYRF 379
Qy 361 DDNSPSTQTSVAKHKKPWVHYIAAEEEDWDYAPLVLAPDDRSYKQYLNNQPORG 420
Db 380 DDNSPSTQTSVAKHKKPWVHYIAAEEEDWDYAPLVLAPDDRSYKQYLNNQPORG 439
Qy 421 RYKVKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTLIIIFKNOASRPYNIYPHGI 480
Db 440 RYKVKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTLIIIFKNOASRPYNIYPHGI 499
Qy 481 TDVRLYRRLPKGVKHLKDFILPGBEIPKYKWTVTVEDGPKSPRCLTRYSSFYVME 540
Db 500 TDVRLYRRLPKGVKHLKDFILPGBEIPKYKWTVTVEDGPKSPRCLTRYSSFYVME 559
Qy 541 RLASGLIGPLLCYKESVDQGNQIMSDKRNVLFSVFDENRSLWLTENIORFLPNPAG 600
Db 560 RLASGLIGPLLCYKESVDQGNQIMSDKRNVLFSVFDENRSLWLTENIORFLPNPAG 619
Qy 601 VOLEDPEFOASIMHISINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTPKH 660
Db 620 VOLEDPEFOASIMHISINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTPKH 679
Qy 661 KMYEDTLTLPFSGETVFMENPGLWILGCHNSDFRNRGHTALLKVSSC---DKWTGD 717
Db 680 KMYEDTLTLPFSGETVFMENPGLWILGCHNSDFRNRGHTALLKVSSCPEGEEDDD 739
Qy 718 YX-----EYSYEDISAYLLSKNAIEPRSF----- 742
Db 740 YLDLEKIFESDDDDYDI-----VDSLIEPRFSQNSRHPSTPQKQFNATTIPENDIEKT 793
Qy 743 ----- 742
Db 794 DPWFARHTPMPKIQNVSSDILLMLLRQSPTHGLSLSLQEAKEYTFSDPSPGADSN 853
Qy 743 ----- 742
Db 854 SLSEMTFRPOLHSGDMVTPEBGLQLRLNKLGTADPLAWNHYGTQIPKEWKSQE 913
Qy 743 -----SONPPVLK 750
Db 914 KSPKTAFFKKOTILSNACESHATAANEQONKEPFIETWAKQORTERLCSQNPVUK 973
Qy 751 RHOREITRTLOSDEIDYDDTISVEMKKEDFDIYDENOSPRSFOKKTRHYFIAAYE 810
Db 974 RHOREITRTLOSDEIDYDDTISVEMKKEDFDIYDENOSPRSFOKKTRHYFIAAYE 1033
Qy 811 RLWDYQWSSPHVLRNRAQSGSVPOPKVVFQBFQFTGSGFTQPLYRGELNEHLGLLPYIR 870
Db 1034 RLWDYQWSSPHVLRNRAQSGSVPOPKVVFQBFQFTGSGFTQPLYRGELNEHLGLLPYIR 1093
Qy 871 AEVEDNIMVTRNQASRPYSFSSLSIYEDDQGAEPKRNFKVKNETKTYFKWQHHMA 930
Db 1094 AEVEDNIMVTRNQASRPYSFSSLSIYEDDQGAEPKRNFKVKNETKTYFKWQHHMA 1153
Qy 931 PTKDEFCKAWAYFSDVLEKDVHSGILIGPLLYCHTNTLNPAGRQVTVQEPALFTTFD 990
Db 1154 PTKDEFCKAWAYFSDVLEKDVHSGILIGPLLYCHTNTLNPAGRQVTVQEPALFTTFD 1213
Qy 991 ETKSWYFTENMERNCRAPCMQIOMEDPTFKENYRFAHNGYIMDTLPLGLVMAQORIRWYL 1050
Db 1214 ETKSWYFTENMERNCRAPCMQIOMEDPTFKENYRFAHNGYIMDTLPLGLVMAQORIRWYL 1273
Qy 1051 LSWGNSNETHSIHFSGHVFTVRKKEVKMALYNLPGVFTVEMLPKAGIWRVECLIGE 1110
Db 1274 LSWGNSNETHSIHFSGHVFTVRKKEVKMALYNLPGVFTVEMLPKAGIWRVECLIGE 1333
Qy 1111 HLHAGMSTLFLVYSNCKQTPGLMASGHIRDFQITASQYQWAPKARLHYSGINAWST 1170
Db 1334 HLHAGMSTLFLVYSNCKQTPGLMASGHIRDFQITASQYQWAPKARLHYSGINAWST 1393
Qy 1171 KEPSWIKVDLLAPMIHIGIKTQAGOKFSSLYISQFIIMYSLDGKKQWYRGNSTGTLM 1230

Db 1394 KEPSWIKVDLLAPMIHIGIKTQAGOKFSSLYISQFIIMYSLDGKKQWYRGNSTGTLM 1453
Qy 1231 VFGNVDSSGIGKNI FNPPIIARIYIRLHPHTYSTRIRLMELMGCDLNSCNPGLMESKA 1290
Db 1454 VFGNVDSSGIGKNI FNPPIIARIYIRLHPHTYSTRIRLMELMGCDLNSCNPGLMESKA 1513
Qy 1291 ISDAQITASSYFTWMEATWSPSKARLHIQGSNARWPOVNNPKWLOVDFOKTMKVTVGT 1350
Db 1514 ISDAQITASSYFTWMEATWSPSKARLHIQGSNARWPOVNNPKWLOVDFOKTMKVTVGT 1573
Qy 1351 TQGVKSILITSMYVKEFLISSQDGHQWTLFFQNGKVKVFOGNQDSFPPVNSLDPPLLTR 1410
Db 1574 TQGVKSILITSMYVKEFLISSQDGHQWTLFFQNGKVKVFOGNQDSFPPVNSLDPPLLTR 1633
Qy 1411 YLRHPQSWHQAIALRMEVLGCEAQDLY 1438
Db 1634 YLRHPQSWHQAIALRMEVLGCEAQDLY 1661

RESULT 4

US-08-558-107-2
; Sequence 2, Application US/08558107
; Patent No. 5910481
; GENERAL INFORMATION:
; APPLICANT: VOORBERG, Johannes J.
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/558,107
; FILING DATE: 13-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ISACSON, John P.
; REGISTRATION NUMBER: 33,715
; REFERENCE/DOCKET NUMBER: 30472/212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1661 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-558-107-2

Query Match 96.7%; Score 7434; DB 2; Length 1661;
Best Local Similarity 86.0%; Pred. No. 0;
Matches 1418; Conservative 4; Mismatches 10; Indels 216; Gaps 4;

Qy 1 ATRRYLGAVALSWDYMQSDLGELPVDARFPFRPKSPFPNTSVVYKKTLPVEFTVHLFN 60
Db 20 ATRRYLGAVALSWDYMQSDLGELPVDARFPFRPKSPFPNTSVVYKKTLPVEFTVHLFN 79
Qy 61 IAKPRPPWMLLGPTTQAQVYDVTVTTLKNMASHPVSLHAGVSVYKASGEGAYDDQTSQ 120
Db 80 IAKPRPPWMLLGPTTQAQVYDVTVTTLKNMASHPVSLHAGVSVYKASGEGAYDDQTSQ 139
Qy 121 REKEDDKVPGGSHYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180

Db 140 REKEDDKVPGSGHTYVQVWLVKNGPMSDPLCLTYSYLSHVLDLVKDLNSGLLGALLVCR 199
Qy 181 EGS LAKEKTQTLHKFTILLFAVFDGKSWHSETKNSLMQDRDAASARAWPMCHTVNGYVNR 240
Db 200 EGS LAKEKTQTLHKFTILLFAVFDGKSWHSETKNSLMQDRDAASARAWPMCHTVNGYVNR 259
Qy 241 SLPLGIGCHRKSYVWIVGHTTPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 300
Db 260 SLPLGIGCHRKSYVWIVGHTTPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 319
Qy 301 MDLQGFLLFCHISSHQHDGNEAVKVDSCPEEPQLMKNEAEEDVDDDLTDEMDVVR 360
Db 320 MDLQGFLLFCHISSHQHDGNEAVKVDSCPEEPQLMKNEAEEDVDDDLTDEMDVVR 379
Qy 361 DDNSPSFIQIRSAKVKHPTWVHYIAAEEEDWDYAPLVLPDDRYSKYQYLANGPQRI 420
Db 380 DDNSPSFIQIRSAKVKHPTWVHYIAAEEEDWDYAPLVLPDDRYSKYQYLANGPQRI 439
Qy 421 RYKVKRPMAYTDETFKTREAIQHESGILGPLLYGEVGTLLIIIFKNQASRPNIYPHGI 480
Db 440 RYKVKRPMAYTDETFKTREAIQHESGILGPLLYGEVGTLLIIIFKNQASRPNIYPHGI 499
Qy 481 TDVRPLYSRRLPKGVKHLKDPFPLPGEIFPKYKWTVTVEDGPTKSDPRCLTRYTSSFWNME 540
Db 500 TDVRPLYSRRLPKGVKHLKDPFPLPGEIFPKYKWTVTVEDGPTKSDPRCLTRYTSSFWNME 559
Qy 541 RDLASGLIGPLLI CYKESVDQRNQIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 600
Db 560 RDLASGLIGPLLI CYKESVDQRNQIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 619
Qy 601 VQLEDFEQASNTMHSINGVYVPSLSQLSVCLHEVAYWYILSIGAQTDFLSVPFSGYTFKH 660
Db 620 VQLEDFEQASNTMHSINGVYVPSLSQLSVCLHEVAYWYILSIGAQTDFLSVPFSGYTFKH 679
Qy 661 KMVYEDTLTPFSGEIVFMSMNPGLWILGCHNSDFRNGMTALLKVSSC---DKNTGD 717
Db 680 KMVYEDTLTPFSGEIVFMSMNPGLWILGCHNSDFRNGMTALLKVSSCIEPEGREDD 739
Qy 718 YY-----EDSYEDISAYLLSKNNAIEPRSF----- 742
Db 740 YLDEKIFSEDDYIDI-----VDSLIEPRFSQNSRHPSTKQKPNATTIPENDIEKT 793
Qy 743 ----- 742
Db 794 DPWFARHTPMPKIQNVSSDLMLLRQSPHGLSLDLQEAKEYETPSDDPSCAIDSN 853
Qy 743 ----- 742
Db 854 SLSEMTFRPOLHSGDMVTPSPGLQLRLNEKLGTTADPLANDNHYGTQIPKEEWSQE 913
Qy 743 -----SONPPVLK 750
Db 914 KSPEKTAFKKOTILSLNACSNHAIAAINEGQNKPEIEVTWAKQGRTERLCSONPPVLK 973
Qy 751 RHOREITRITLQSDQEBIDYDDTISVEMKEDFDIYDENENQSPRSQKTRHYFFIAAVE 810
Db 974 RHOREITRITLQSDQEBIDYDDTISVEMKEDFDIYDENENQSPRSQKTRHYFFIAAVE 1033
Qy 811 RLMDYGNSSPHVLRNRAQSGVPOFKKVYQEBTDCGFTQPLYRGELNHLGLLGPYIR 870
Db 1034 RLMDYGNSSPHVLRNRAQSGVPOFKKVYQEBTDCGFTQPLYRGELNHLGLLGPYIR 1093
Qy 871 AEVEDNTMTFRNQASRPYFYSLSIYSEEDQRCQABPRKNFVKPNETKTYFWKVQHMA 930
Db 1094 AEVEDNTMTFRNQASRPYFYSLSIYSEEDQRCQABPRKNFVKPNETKTYFWKVQHMA 1153
Qy 931 PTKDEDFCKAWAYSDVDLEKOVHSGLIGPLLVCHTNTLPAHGRQYTVQEFALFFTFID 990
Db 1154 PTKDEDFCKAWAYSDVDLEKOVHSGLIGPLLVCHTNTLPAHGRQYTVQEFALFFTFID 1213
Qy 991 ETKSWYFTENMERNCRAPCNIQMEDPTFKENYRPHAINGYIMDTPLGLVMAQDQIRWYL 1050
Db 1214 ETKSWYFTENMERNCRAPCNIQMEDPTFKENYRPHAINGYIMDTPLGLVMAQDQIRWYL 1273

Qy 1051 LMSGNENIHSIHFGSHVFTVRKCEBYKQALINLIFGVFTVEMLPSPKAGIWRVECLIGE 1110
Db 1274 LMSGNENIHSIHFGSHVFTVRKCEBYKQALINLIFGVFTVEMLPSPKAGIWRVECLIGE 1333
Qy 1111 HLHAGMSTLFLVYSNKKQTPLGWASGHIRDFQITASGOYGQWAPKLARLHYSGSINAWST 1170
Db 1334 HLHAGMSTLFLVYSNKKQTPLGWASGHIRDFQITASGOYGQWAPKLARLHYSGSINAWST 1393
Qy 1171 KEPPFSWIKVDDLAPMIHIGIKTQGARQKFFSSLYISOFIIMYSLDGKKWQTYRGNSTGTLM 1230
Db 1394 KEPPFSWIKVDDLAPMIHIGIKTQGARQKFFSSLYISOFIIMYSLDGKKWQTYRGNSTGTLM 1453
Qy 1231 VFQGNVDSGIGIKENIFNPPIIARIYRLHPTHYSIRSLRMELMGCDLNSCMLPGMESKA 1290
Db 1454 VFQGNVDSGIGIKENIFNPPIIARIYRLHPTHYSIRSLRMELMGCDLNSCMLPGMESKA 1513
Qy 1291 ISDAQITASSYFTNMFATWSPSKARLHLQGRSNAMRPVNNPKEWLQVDFQKTMKVTGVT 1350
Db 1514 ISDAQITASSYFTNMFATWSPSKARLHLQGRSNAMRPVNNPKEWLQVDFQKTMKVTGVT 1573
Qy 1351 TQGVKSLLTSMYKKEFLISSQDGHQWTLFFQNGKVKVFGQNDSTFPVNSLDPELLTR 1410
Db 1574 TQGVKSLLTSMYKKEFLISSQDGHQWTLFFQNGKVKVFGQNDSTFPVNSLDPELLTR 1633
Qy 1411 YLRHQPQSWHQIALRMEVLGCEAQDLY 1438
Db 1634 YLRHQPQSWHQIALRMEVLGCEAQDLY 1661

RESULT 5

US-09-243-539-2
Sequence 2, Application US/09243539
Patent No. 6130203
GENERAL INFORMATION:
APPLICANT: VOORBERG, Johannes J.
TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/243,539
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/558,107
FILING DATE: 13-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: ISACSON, John P.
REGISTRATION NUMBER: 33,715
REFERENCE/DOCKET NUMBER: 30472/212
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1661 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-243-539-2

Query Match 96.78; Score 7434; DB 3; Length 1661;
 Best Local Similarity 86.08; Pred. No. 0;
 Matches 1416; Conservative 4; Mismatches 10; Indels 216; Gaps 4;

QY 1 ATRRYLGAVELSNDYMQSDLGELPVDAREPPRPVPSFPNTSVVYKTKLTFVEFTVHLEN 60
 DB 20 ATRRYLGAVELSNDYMQSDLGELPVDAREPPRPVPSFPNTSVVYKTKLTFVEFTVHLEN 79
 QY 61 IAKPRPMMGLGPTIOAEVYDVTWITLKNMASHPVSLHAGVSYKASEGAEYDDOTSQ 120
 DB 80 IAKPRPMMGLGPTIOAEVYDVTWITLKNMASHPVSLHAGVSYKASEGAEYDDOTSQ 139
 QY 121 REKEDDVFPGGSHYVWVLKENGPMASDPLCLITYSLSHVDLVKDLNSGLIGALLVCR 180
 DB 140 REKEDDVFPGGSHYVWVLKENGPMASDPLCLITYSLSHVDLVKDLNSGLIGALLVCR 199
 QY 181 EGSIAKKTOTLHKPILFPAVDFRGKSHWSETKNSLMQDDAASARAWPMHTVNGYVR 240
 DB 200 EGSIAKKTOTLHKPILFPAVDFRGKSHWSETKNSLMQDDAASARAWPMHTVNGYVR 259
 QY 241 SLPLGLICHRKSVVWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLL 300
 DB 260 SLPLGLICHRKSVVWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLL 319
 QY 301 MDLQOFLFLCHISSHODGMEAYVKVDCPEEPQLRMKNNEAEYDDDLTDSMDVVRP 360
 DB 320 MDLQOFLFLCHISSHODGMEAYVKVDCPEEPQLRMKNNEAEYDDDLTDSMDVVRP 379
 QY 361 DDNSPSFIQIRSAKHKPTWHYIAAEEDWDYAPLVLPDDRYSKYQYLNGPQRI 420
 DB 380 DDNSPSFIQIRSAKHKPTWHYIAAEEDWDYAPLVLPDDRYSKYQYLNGPQRI 439
 QY 421 RYKVKRFMAYTDTFTKTRAIQHESGILGLYGEVGDTLIIIFKNQASRPNIYPHGI 480
 DB 440 RYKVKRFMAYTDTFTKTRAIQHESGILGLYGEVGDTLIIIFKNQASRPNIYPHGI 499
 QY 481 TDVRPLYSRLPKGVKHLKDFPILPGEIFKYKMTVTVEDGPTKSDPRCLTRYYSFVNME 540
 DB 500 TDVRPLYSRLPKGVKHLKDFPILPGEIFKYKMTVTVEDGPTKSDPRCLTRYYSFVNME 559
 QY 541 RDLASGLIGPLIICYKESVDORGNQIMSKRNVLFSVDENESWLTENIQRLFNPA 600
 DB 560 RDLASGLIGPLIICYKESVDORGNQIMSKRNVLFSVDENESWLTENIQRLFNPA 619
 QY 601 VQLEDPFQASNMHSGVVFQSLQSVCLHEVAYWYILSIGAQDTFLSVFSGYTFKH 660
 DB 620 VQLEDPFQASNMHSGVVFQSLQSVCLHEVAYWYILSIGAQDTFLSVFSGYTFKH 679
 QY 661 KMVEDTLTLPPSGETVFMENPGLWILGCHNSDFRNGMTALLKVSSC----DKWTD 717
 DB 680 KMVEDTLTLPPSGETVFMENPGLWILGCHNSDFRNGMTALLKVSSCIPGREDDD 739
 QY 718 YY-----EDSYEDISAYLLSKNAIEPRSF----- 742
 DB 740 YLDLEKIFSEDDYIDI-----VDSLIEPRSFQNSRHPSTKQFNATTIPENDIEKT 793
 QY 743 ----- 742
 DB 794 DPMFAHTRPMKIQNVSSDMLLMLRQSPTPHGLSLASDLQEAKEYTFSDPSPGAIDSN 853
 QY 743 ----- 742
 DB 854 SLSEWTHFRPOLHSGDNVFTPSGLQRLNEKLGTTADPLAWNHYGQIPKEWKSQB 913
 QY 743 -----SONPPVLK 750
 DB 914 KSPEKTAFFKXDTLSLNACESNHAIAINEGQNKPEIEVTWAKQGRTERLCSQNPVLK 973
 QY 751 RHQREITRTTLOSQOEIDYDDTISVEMKEDFDIYDENQSPRQFKTRHYFFIAAVE 810
 DB 974 RHQREITRTTLOSQOEIDYDDTISVEMKEDFDIYDENQSPRQFKTRHYFFIAAVE 1033
 QY 811 RLWDYGNSSSHVLRNRAQSGSVQPKKVVQFEFTDGSFTQPLYRGELNEHLGLGPYIR 870

DB 1034 RLWDYGNSSSHVLRNRAQSGSVQPKKVVQFEFTDGSFTQPLYRGELNEHLGLGPYIR 1093
 QY 871 AEVEDNIMVTRNOASRPYSFYSLISYEBDQRCQABPRKNFKVPKPNSTKTYFKVQHHMA 930
 DB 1094 AEVEDNIMVTRNOASRPYSFYSLISYEBDQRCQABPRKNFKVPKPNSTKTYFKVQHHMA 1153
 QY 931 PTKDEFCKAWAYESDVLEKDVHSGGLIGPLLVCHTNTLPAHGRQVTVQEPALFTI 990
 DB 1154 PTKDEFCKAWAYESDVLEKDVHSGGLIGPLLVCHTNTLPAHGRQVTVQEPALFTI 1213
 QY 991 ETKSWYFTENMERNCRAPCMIQMEDPTFKENYRFAHNGYIMDTLPGLVNAQOQIRWYL 1050
 DB 1214 ETKSWYFTENMERNCRAPCMIQMEDPTFKENYRFAHNGYIMDTLPGLVNAQOQIRWYL 1273
 QY 1051 LSMGSNENIHSIHESGHVFTVRKKEEYKMAIYNYLPQVFTVEMLPKAGIWRVECLIGE 1110
 DB 1274 LSMGSNENIHSIHESGHVFTVRKKEEYKMAIYNYLPQVFTVEMLPKAGIWRVECLIGE 1333
 QY 1111 HLHAGMSTLFLVYSNKCQTPLGMAHGHIRDFQITASQYQWAPKLARLHYSGINAWST 1170
 DB 1334 HLHAGMSTLFLVYSNKCQTPLGMAHGHIRDFQITASQYQWAPKLARLHYSGINAWST 1393
 QY 1171 KEPPSWIKVDLLAPMIHIGIKTQGAROKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLM 1230
 DB 1394 KEPPSWIKVDLLAPMIHIGIKTQGAROKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLM 1453
 QY 1231 VFFGNVDSSGGLKHNIFNPPIIARVIRLHPHTYSTRMLMELMGLCDLNSCMLPGMESKA 1290
 DB 1454 VFFGNVDSSGGLKHNIFNPPIIARVIRLHPHTYSTRMLMELMGLCDLNSCMLPGMESKA 1513
 QY 1291 ISDAQITASSYFTNMFTWSPSKARLHQGRSNARQVANNPKWQLQVDFQTKMKVTGVT 1350
 DB 1514 ISDAQITASSYFTNMFTWSPSKARLHQGRSNARQVANNPKWQLQVDFQTKMKVTGVT 1573
 QY 1351 TQGVKSLTSMYVKEFLISSQDGHQWTLFPQNGKVKVFGQNOQSDTFPVVNSLDPPLLTR 1410
 DB 1574 TQGVKSLTSMYVKEFLISSQDGHQWTLFPQNGKVKVFGQNOQSDTFPVVNSLDPPLLTR 1633
 QY 1411 YLRHPQSWHQAIALRMEVLGCEADLY 1438
 DB 1634 YLRHPQSWHQAIALRMEVLGCEADLY 1661

RESULT 6
 US-07-864-004B-4
 ; Sequence 4, Application US/07864004B
 ; Patent No. 5364771
 ; GENERAL INFORMATION:
 ; APPLICANT: Lollar, John S.
 ; APPLICANT: Runge, Marshall S.
 ; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Kilpatrick & Cody
 ; STREET: 1100 Peachtree Street
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: US
 ; ZIP: 30309
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/864,004B
 ; FILING DATE: 07 APRIL 1992
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Pabst, Patrea L.
 ; REGISTRATION NUMBER: 31,284
 ; REFERENCE/DOCKET NUMBER: EMU106

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-815-6508

TELEFAX: 404-815-6555

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 2332 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO N-terminal

FRAGMENT TYPE: N-terminal

ORGANISM: Homo sapien

TISSUE TYPE: Liver cDNA sequence

US-07-864-004B-4

Query Match 94.1%; Score 7234; DB 1; Length 2332;

Best Local Similarity 61.7%; Pred. No. 0;

Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;

```
QY 1 ATRRYLGVAVELSDYQSDLGELPVDAREPPRPVPSKSPFPPNTSVVYKKTLLFVEFTVHLFN 60
DB 1 ATRRYLGVAVELSDYQSDLGELPVDAREPPRPVPSKSPFPPNTSVVYKKTLLFVEFTVHLFN 60
QY 61 IAKPRPPWMLGGLTQIAEYDVTWITLKNMASHPVSLHAGVSYWKASGEAYDDQTSQ 120
DB 61 IAKPRPPWMLGGLTQIAEYDVTWITLKNMASHPVSLHAGVSYWKASGEAYDDQTSQ 120
QY 121 REKEDDKVPFGGSHTYVQVLKNGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
DB 121 REKEDDKVPFGGSHTYVQVLKNGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 181 EGSILAKETQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKQHTVNGYVNR 240
DB 181 EGSILAKETQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKQHTVNGYVNR 240
QY 241 SLFGLIGCHRKSVYHVGITGTTPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 300
DB 241 SLFGLIGCHRKSVYHVGITGTTPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 300
QY 301 MDLQGFLLFCHISSHQHDGMEAYVKVDSCEEPQLRMKNNEAEYDDDLTDSMDVVRP 360
DB 301 MDLQGFLLFCHISSHQHDGMEAYVKVDSCEEPQLRMKNNEAEYDDDLTDSMDVVRP 360
QY 361 DDNSPSFIQIRSVAKKPKTWVHYIAAREEDWDYAPLVLAPODRSVKSOYIANGPQRI 420
DB 361 DDNSPSFIQIRSVAKKPKTWVHYIAAREEDWDYAPLVLAPODRSVKSOYIANGPQRI 420
QY 421 RYKVKVRFMAYTDETFKTREAIQHSIGILGPLLYGEVGTLLIIFKNQASRPYNIYPHGI 480
DB 421 RYKVKVRFMAYTDETFKTREAIQHSIGILGPLLYGEVGTLLIIFKNQASRPYNIYPHGI 480
QY 481 TDVRLYLRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGPTKSDPCLFRYYSFVNME 540
DB 481 TDVRLYLRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGPTKSDPCLFRYYSFVNME 540
QY 541 RDLASGLIGPLLIICYKESVDQNGQIMSDGNVILFSVPDENRSWLTENIQRLPNPAG 600
DB 541 RDLASGLIGPLLIICYKESVDQNGQIMSDGNVILFSVPDENRSWLTENIQRLPNPAG 600
QY 601 VQLEDPEFQASNIHMSINGVFDLSQLSVCLHEVAYWYIISIGAQTDPLSVFSGGYTFKH 660
DB 601 VQLEDPEFQASNIHMSINGVFDLSQLSVCLHEVAYWYIISIGAQTDPLSVFSGGYTFKH 660
QY 661 KMVYEDTLTLFPFSGEYTFVMSMENPGLWILGCHNSDFRNRMGTALLKVSSCDKNTGDYYE 720
DB 661 KMVYEDTLTLFPFSGEYTFVMSMENPGLWILGCHNSDFRNRMGTALLKVSSCDKNTGDYYE 720
QY 721 DSVEDISAYLLSKNNAIEPRSF----- 742
DB 721 DSVEDISAYLLSKNNAIEPRSF----- 742
```

967 NTLPAGHQVTVQEFALFFITFDETSKSWYFTEMNCRAPCNIMQEDPTFKENYRPHA 1026
1861 NTLPAGHQVTVQEFALFFITFDETSKSWYFTEMNCRAPCNIMQEDPTFKENYRPHA 1920
1027 INGIMDTLPGVMAQDQIRWYLLSMGNSNENIHSIHPSGHVFTVRKKEEKMALYNLYP 1086
1921 INGIMDTLPGVMAQDQIRWYLLSMGNSNENIHSIHPSGHVFTVRKKEEKMALYNLYP 1980
1087 GVFEVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNCKOTPLGMAHGHIRPQITAS 1146
1981 GVFEVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNCKOTPLGMAHGHIRPQITAS 2040
1147 GOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQARQKFSLSYISQ 1206
2041 GOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQARQKFSLSYISQ 2100
1207 FIIMYSLDGKKQWTVRGNGSTGTLMVFFGNVDSGIGKNI FNPPIIARIYIRLHPHTYSIRS 1266
2101 FIIMYSLDGKKQWTVRGNGSTGTLMVFFGNVDSGIGKNI FNPPIIARIYIRLHPHTYSIRS 2160
1267 TLRWELMGCDLNSCMPGIESKALSDAQITASSYFTNMFTWSPSKARLHLQGRSNAWR 1326
2161 TLRWELMGCDLNSCMPGIESKALSDAQITASSYFTNMFTWSPSKARLHLQGRSNAWR 2220
1327 PQVNNPKWLQVDFOKTMKVTVGTTQGVKSLTSMYKFEFLISSQDGHQWTLFPQNGKV 1386
2221 PQVNNPKWLQVDFOKTMKVTVGTTQGVKSLTSMYKFEFLISSQDGHQWTLFPQNGKV 2280
1387 KVFQGNQSFPTVNVNSLDPPLITRYLRHPQSWVHQIALRMEVLGCEAODLY 1438
2281 KVFQGNQSFPTVNVNSLDPPLITRYLRHPQSWVHQIALRMEVLGCEAODLY 2332

RESULT 7

US-08-251-937A-4
Sequence 4, Application US/08251937A
Patent No. 5583209
GENERAL INFORMATION:
APPLICANT: Lollar, John S.
APPLICANT: Runge, Marshall S.
TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 1100 Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/251,937A
FILING DATE: 31-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pratt, John S.
REGISTRATION NUMBER: 29,476
REFERENCE/DOCKET NUMBER: EMU106D1V
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-815-6367
TELEFAX: 404-815-6555
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapien
TISSUE TYPE: Liver cdna sequence
US-08-251-937A-4
Query Match 94.1%; Score 7234; DB 1; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
QY 1 ATRYIYLGAVELSDWYMQSDLGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTVHLFN 60
DB 1 ATRYIYLGAVELSDWYMQSDLGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTVHLFN 60
QY 61 IAKPRPPWMLGGLGPTIOAEVYDVTVIITLKNMASHPVSLHAGVSYWKASGAEYDDQTSQ 120
DB 61 IAKPRPPWMLGGLGPTIOAEVYDVTVIITLKNMASHPVSLHAGVSYWKASGAEYDDQTSQ 120
QY 121 REKEDDKVFPGGSHYVQVLEKNGPMASDPLCLITYSLSHVDLVKDINSGLIGALLVCR 180
DB 121 REKEDDKVFPGGSHYVQVLEKNGPMASDPLCLITYSLSHVDLVKDINSGLIGALLVCR 180
QY 181 EGSIAKKTOTLHKPFIILFAVFDGKGNHSTKNSLMODRDAASARAWPKMHTVNGYVNR 240
DB 181 EGSIAKKTOTLHKPFIILFAVFDGKGNHSTKNSLMODRDAASARAWPKMHTVNGYVNR 240
QY 241 SLPLGIGCHRSVYWHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPIITLTAQLL 300
DB 241 SLPLGIGCHRSVYWHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPIITLTAQLL 300
QY 301 MDLQGFLLFCHSSHQHDGMEAYVKVDSCEPEPQLRMKNNEAEYDDDLTDSMDVYVRF 360
DB 301 MDLQGFLLFCHSSHQHDGMEAYVKVDSCEPEPQLRMKNNEAEYDDDLTDSMDVYVRF 360
QY 361 DDNSPSFIQIRSVAKKHPTKWTWVHYIAAEEDWDYAPLVLPADDDRSYKQYLNNPQIRG 420
DB 361 DDNSPSFIQIRSVAKKHPTKWTWVHYIAAEEDWDYAPLVLPADDDRSYKQYLNNPQIRG 420
QY 421 RYKVKRPMAYTDETFKTREAIQHSIGILGPLYGEVGDTLIIIFKNQASRYNIIYPHGI 480
DB 421 RYKVKRPMAYTDETFKTREAIQHSIGILGPLYGEVGDTLIIIFKNQASRYNIIYPHGI 480
QY 481 TDVRLYSRRLPKGVKHLKDFILPGEIFKVKWTVTVEDGPTKSDPRCLTRYSSFVNM 540
DB 481 TDVRLYSRRLPKGVKHLKDFILPGEIFKVKWTVTVEDGPTKSDPRCLTRYSSFVNM 540
QY 541 RDLASGLIGPLLCYKESVDQGNQIMSDKRNVLIFSVPDENRSWYLTENIORFLPNPAG 600
DB 541 RDLASGLIGPLLCYKESVDQGNQIMSDKRNVLIFSVPDENRSWYLTENIORFLPNPAG 600
QY 601 VQLEDPFQASNIMHSINGYVDSIQLSVCLHEVAYWYTLISGATDFLVSFFSGYTFKH 660
DB 601 VQLEDPFQASNIMHSINGYVDSIQLSVCLHEVAYWYTLISGATDFLVSFFSGYTFKH 660
QY 661 KMVYEDTLTLFPFSGETVFMSENGLWITLGNCHSDFRNRGMTALLIKVSSCDKNTGDYYE 720
DB 661 KMVYEDTLTLFPFSGETVFMSENGLWITLGNCHSDFRNRGMTALLIKVSSCDKNTGDYYE 720
QY 721 DSYEDISAYLLSKNNAIEPRSF-----742
DB 721 DSYEDISAYLLSKNNAIEPRSFQNSRHPSTQKQFNATTIPENDIEKTDPPFAHRTMP 780
QY 743 -----742
DB 781 KIQNVSSDILLMLLRQSPTPHGLSLDLOEAKYETFSDDPSGAIIDNNLSLSEMTFRPQ 840
QY 743 -----742
DB 841 LHHSGUMVTPRSGIQLRLNEKLGTTAATLKKLDFKVSSTSNLLISTIPSDNLAAGTDN 900

QY 743 ----- 742
Db 901 TSSLGPPMPVHYDSQLDTTLFGKSSPLTESGGPLSLSENNDSKLESGLMNSQESSW 960
QY 743 ----- 742
Db 961 GKNVSTESGRFLFKGRAHPALLTKONALFKVSIISLLTKNTKTSNNSATNRKTHIDGSL 1020
QY 743 ----- 742
Db 1021 LIENSPVQNIILESTDEPKVTPLIHDRMLMDKNATALRLNHSNKTTSKNWEMVQK 1080
QY 743 ----- 742
Db 1081 KEGPIPPAQNDPMSFFKMLFLPESARWIQRTHGKNSLNSGGQSPKQLVSLGPEKSVEG 1140
QY 743 ----- 742
Db 1141 QNFLSEKNVVGKGEFTKDVGLKEMVFPSSRLFTNLNLHNNTHNQEKIQBIEK 1200
QY 743 ----- 742
Db 1201 KETLIQENVVLPOIHVTVGTNFKMKNLFLLSRQNVESYEGAYAPVLQDFRSLNDSTNR 1260
QY 743 ----- 742
Db 1261 TKKHTAFSKGBEENLEGLGNQTKOIVEKYACTTRISPNTSQNFVQSRKRALQFRL 1320
QY 743 ----- 742
Db 1321 PLEETELEKRIIVDDTSTOWSKNMKHLTPSLTAQIDYNEKEGAITQSPSLDCLTRSHI 1380
QY 743 ----- 742
Db 1381 PQANRSPPLAKVSSPFSIRPIVLRVLFQDNSSHLPAASYRKDSGVQESSHFLQAKK 1440
QY 743 ----- 742
Db 1441 NNLSLAILTEMTGDQREVSLGTSATNSVYKKVENTVLPKPDLPKTSKGVKELLPKVHI 1500
QY 743 ----- 742
Db 1501 YQKDLPTTETSGPHLDLVEGSLLOGTEGAIKWNEANRPGKVPFLRVATESSAKTPSK 1560
QY 743 ----- 742
Db 1561 LLDPLAWDNHYGTQIPKEEWSQESKPEKTAFFKKDTILSLNACESHATAAINEGONKP 1620
QY 743 -----SQNPVLKRHQREITRITLQSDQBEIDYDDTISVEMKEDFDIY 786
Db 1621 EIEVTWAKQGRTERLCSQNPVLKRHQREITRITLQSDQBEIDYDDTISVEMKEDFDIY 1680
QY 787 DEENQSPRSFQKTRHYFIAAVERLWDYQWSSSPHVLNRNAGSGVPQKVVQFQFTD 846
Db 1681 DEENQSPRSFQKTRHYFIAAVERLWDYQWSSSPHVLNRNAGSGVPQKVVQFQFTD 1740
QY 847 GSFTQPLYRGELNEHLGLGPYRAEVEDNIMVTFRQASRPYSFYSSLLISYEDQOQA 906
Db 1741 GSFTQPLYRGELNEHLGLGPYRAEVEDNIMVTFRQASRPYSFYSSLLISYEDQOQA 1800
QY 907 EPRKNFKVNETKTYFKWQHQHMAPTDFDCKAWYFSDVLEKDVHSGLIGPLVCHT 966
Db 1801 EPRKNFKVNETKTYFKWQHQHMAPTDFDCKAWYFSDVLEKDVHSGLIGPLVCHT 1860
QY 967 NTLNPAHGRQVTVQEFALFTTIFDETSKYFTENMERNCRAPCNQIMEDPTFKENYRPHA 1026
Db 1861 NTLNPAHGRQVTVQEFALFTTIFDETSKYFTENMERNCRAPCNQIMEDPTFKENYRPHA 1920
QY 1027 INGYIMDTLPLGLVMAQORTRWILLSMGSENTHSFSGHVFTVRKKEBKALYNLYP 1086
Db 1921 INGYIMDTLPLGLVMAQORTRWILLSMGSENTHSFSGHVFTVRKKEBKALYNLYP 1980

QY 1087 GYFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGMASGHIRDFOITAS 1146
Db 1981 GYFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGMASGHIRDFOITAS 2040
QY 1147 GQYGWAPKLARLHYSGSINAMSTKEPFSWIKVDLLAPMIHGIKTQGARQKFSSLYISQ 1206
Db 2041 GQYGWAPKLARLHYSGSINAMSTKEPFSWIKVDLLAPMIHGIKTQGARQKFSSLYISQ 2100
QY 1207 FIIMYSLDGKKQVOTYRGNSTGTLMVFFGNVDSSGKHNIFNPPIIARYIRLHPTHYSIRS 1266
Db 2101 FIIMYSLDGKKQVOTYRGNSTGTLMVFFGNVDSSGKHNIFNPPIIARYIRLHPTHYSIRS 2160
QY 1267 TLRMELMGCDLNSCMLGMEKSAISDAQITASSYFTNMFATWSPSKARLHLQGRSNWR 1326
Db 2161 TLRMELMGCDLNSCMLGMEKSAISDAQITASSYFTNMFATWSPSKARLHLQGRSNWR 2220
QY 1327 PQVNNPKELQVDFQKTMKVTTQGVKSLLTSMYVKEFLISSSDGHWTLFFQNGKV 1386
Db 2221 PQVNNPKELQVDFQKTMKVTTQGVKSLLTSMYVKEFLISSSDGHWTLFFQNGKV 2280
QY 1387 KVFQGNQDSFTPVNSLDPPLTRYLRIRHQPQSVHQAIALRMEVLCGEADLY 1438
Db 2281 KVFQGNQDSFTPVNSLDPPLTRYLRIRHQPQSVHQAIALRMEVLCGEADLY 2332

RESULT 8

US-08-212-133A-2
; Sequence 2, Application US/08212133A
; Patent No. 5663060
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Runge, Marshall S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212,133A
; FILING DATE: March 11, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Padst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU/76677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-572-6508
; TELEFAX: 404-572-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence

US-08-212-133A-2

Query Match				94.1%;	Score 7234;	DB 1;	Length 2332;
Best Local Similarity				61.7%;	Pred. No. 0;		
Matches 1438;				Conservative	0;	Mismatches	0;
				Indels	894;	Gaps	1;
QY	1	ATRYIYGAVELSDYMQSDLGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTVHLFN	60				
DB	1	ATRYIYGAVELSDYMQSDLGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTVHLFN	60				
QY	61	IAPRRPMMGLLGTPTIAQEVYDVTIVITLKNASHPVSLHVGVSYNKASGAEVDDQTSQ	120				
DB	61	IAPRRPMMGLLGTPTIAQEVYDVTIVITLKNASHPVSLHVGVSYNKASGAEVDDQTSQ	120				
QY	121	REKEDDKVFPGGSHYVQVWIKENGPMASDPLCLITYSLSHVDLVKDLNSGLIGALLVCR	180				
DB	121	REKEDDKVFPGGSHYVQVWIKENGPMASDPLCLITYSLSHVDLVKDLNSGLIGALLVCR	180				
QY	181	EGSLAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR	240				
DB	181	EGSLAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR	240				
QY	241	SLPGLIGCHRKSVMVHVGMTTPEVHSIFLEGHTFLVRNROASLEISPIFTLTAQTL	300				
DB	241	SLPGLIGCHRKSVMVHVGMTTPEVHSIFLEGHTFLVRNROASLEISPIFTLTAQTL	300				
QY	301	MDLQGFLLFCHISSHQDHGMEAYVKVDSCPEEPQLRMKNNEAEYDDDLTDSEMDVVRP	360				
DB	301	MDLQGFLLFCHISSHQDHGMEAYVKVDSCPEEPQLRMKNNEAEYDDDLTDSEMDVVRP	360				
QY	361	DDNSPSPSIQIRSAKHPKTVHVIYAAEEDWDYAPLVLAAPDDRSYKSYQLNNGPQIRG	420				
DB	361	DDNSPSPSIQIRSAKHPKTVHVIYAAEEDWDYAPLVLAAPDDRSYKSYQLNNGPQIRG	420				
QY	421	RYKVKVRMAYTDTFTKTRAIQHSGLIGPLLYGCVGDTLLIIFKNOASRPYNIYPKGI	480				
DB	421	RYKVKVRMAYTDTFTKTRAIQHSGLIGPLLYGCVGDTLLIIFKNOASRPYNIYPKGI	480				
QY	481	TDVRLYSRRLPKGVKHLKDPILPGEIFKYKWTVTVEDGPKSPRCLTRYSSFVNME	540				
DB	481	TDVRLYSRRLPKGVKHLKDPILPGEIFKYKWTVTVEDGPKSPRCLTRYSSFVNME	540				
QY	541	RDASGLIGPLLI CYKESVDQGNQIMSDKRNVLIFSVDENRSLYLTENIQRLFPNAG	600				
DB	541	RDASGLIGPLLI CYKESVDQGNQIMSDKRNVLIFSVDENRSLYLTENIQRLFPNAG	600				
QY	601	VOLEDEPOASIMHSINGYVDFSLSVCLHEVAYWYILSGAOTDFLSVFFSGTTFKH	660				
DB	601	VOLEDEPOASIMHSINGYVDFSLSVCLHEVAYWYILSGAOTDFLSVFFSGTTFKH	660				
QY	661	KMYEDTLTLFPFGSGTVMFMSNPGMLWILGCHNSDFRNGMTALLKVSSCDKNTGDYYE	720				
DB	661	KMYEDTLTLFPFGSGTVMFMSNPGMLWILGCHNSDFRNGMTALLKVSSCDKNTGDYYE	720				
QY	721	DSYEDISAYLLSKNNAIEPRSF	742				
DB	721	DSYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKFNATTIPENDIEKTPMFAHRTMP	780				
QY	743		742				
DB	781	KIQNVSSDILLMLLRQSPTHGLSLDQEAKEYTFDDPPSGAIDSNNSLSEMTFRPQ	840				
QY	743		742				
DB	841	LHSGDMVFTPEGLQLRLNEKLGTTAATLKLDPKVSSTSNLITIPSDNLAAGTDN	900				
QY	743		742				
DB	901	TSSIGPSPMPHYDSQDITTLFGKSSPLTESGGPLSLSBENNDKLLSGLMNSQESSW	960				
QY	743		742				
DB	961	GKNVSTESGRLFKGRAHPALITKDNALFKVSIILLKTNKTSNNSATNRKTHIDGPSL	1020				

QY	743	-----	742
DB	1021	LIENSPVMQNIILESTEFKVTPLIHDRMLMDKNATLRLNMSNKTTSKKNMVAQK	1080
QY	743	-----	742
DB	1081	KEGPIDPAQNDPMSFFKMLFLPESARWIQRTGKNLSNGQSPSKQLVSLGPEKSV	1140
QY	743	-----	742
DB	1141	QNFLSEKNVVGKGEFTKDVGLKEMVFPSSRNLFITNLDNLHNNTHNQEKIOBEIEK	1200
QY	743	-----	742
DB	1201	KETLIOENVVLPIQIHTVTGTFKMFKNLFLSTRQNVESYEGAYAPVLQFRSLNDSTNR	1260
QY	743	-----	742
DB	1261	TKKHTAHFSKKGEHENLEGLGNQTKQIVEKVACTTRISPNTSQNFVQSRKRALQKPL	1320
QY	743	-----	742
DB	1321	PLBETELEKRIIVDDTSTQSKNMKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHI	1380
QY	743	-----	742
DB	1381	PQANRSLPIAKVSSFPSPSIRPIYLRVLFDONSHLPAASRYRKDQSGVQSSHFLOGAK	1440
QY	743	-----	742
DB	1441	NNLSLAITLMTQDQREVGSLGTSATNSVYKKVENTVLPKPLPKTSGKVELLPKVI	1500
QY	743	-----	742
DB	1501	YQDLFPETTSNGSPGHLDLVEGSLQTEGAIKWNEANRPKVFLRVATESSAKTPSK	1560
QY	743	-----	742
DB	1561	LLDPLANDHNYGTQIPKEWKSQKSPKTAFFKDDTILSNACESNHAIAINEGQNK	1620
QY	743	-----SONPPVLKRHQREIITRTTLQSDQBEIDYDDTISVEMKKEDFDIY	786
DB	1621	EIEVTNAKQGTERTLCSONPPVLKRHQREIITRTTLQSDQBEIDYDDTISVEMKKEDFDIY	1680
QY	787	DEDENQSPRSQKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGSVQFKKVFQETD	846
DB	1681	DEDENQSPRSQKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGSVQFKKVFQETD	1740
QY	847	GSFTQPLRGELNEHGLLGPYIRAEVEDNIMVTFRNOASRPYSFVSSLSIYEEQROGA	906
DB	1741	GSFTQPLRGELNEHGLLGPYIRAEVEDNIMVTFRNOASRPYSFVSSLSIYEEQROGA	1800
QY	907	EPKKNFVKPNETKTYFKVQHHMPTKDFCDKAWAYFSDVDLEKDVHSGGLIGPLLVCHT	966
DB	1801	EPKKNFVKPNETKTYFKVQHHMPTKDFCDKAWAYFSDVDLEKDVHSGGLIGPLLVCHT	1860
QY	967	NTLNPAHQGVTVQBFALFFTFIDETKSWYFTENMERNCRAPCNTOEMDPTFKENYRFA	1026
DB	1861	NTLNPAHQGVTVQBFALFFTFIDETKSWYFTENMERNCRAPCNTOEMDPTFKENYRFA	1920
QY	1027	INGYIMDTLPLGLVMAQDQIRWYLLSMGSNENIHSIFSGHVFTVRKKEEKWALYNLYP	1086
DB	1921	INGYIMDTLPLGLVMAQDQIRWYLLSMGSNENIHSIFSGHVFTVRKKEEKWALYNLYP	1980
QY	1087	GVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKCTPLGMASGHIRDQITAS	1146
DB	1981	GVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKCTPLGMASGHIRDQITAS	2040
QY	1147	GOYGWAPKLARLHSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQAGARQFSSLIYSQ	1206
DB	2041	GOYGWAPKLARLHSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQAGARQFSSLIYSQ	2100

QY 1207 FIIMSLDGKKWQYRGNSTGTLVFPFGNVDSSGKIKNINFPPIIARIIRLHPTHYSIRS 1266
DB 2101 FIIMSLDGKKWQYRGNSTGTLVFPFGNVDSSGKIKNINFPPIIARIIRLHPTHYSIRS 2160
QY 1267 TLAMELWGCNLSNCSPLGMSKASDAQITASSYFTNMFATWSPSKARLHLQGRSNWR 1326
DB 2161 TLAMELWGCNLSNCSPLGMSKASDAQITASSYFTNMFATWSPSKARLHLQGRSNWR 2220
QY 1327 PQVNNPKEMQLQVDFQKTMKVTGVTTQGVKSLLTSMYKBEFLISSQDGHQWTLFFQNGKV 1386
DB 2221 PQVNNPKEMQLQVDFQKTMKVTGVTTQGVKSLLTSMYKBEFLISSQDGHQWTLFFQNGKV 2280
QY 1387 KVFQGNQDSFTPVNLSLDPPLLRYLRIHQSWHQTALRMEVLGCEAODLY 1438
DB 2281 KVFQGNQDSFTPVNLSLDPPLLRYLRIHQSWHQTALRMEVLGCEAODLY 2332

RESULT 9

US-08-474-503-2
; Sequence 2, Application US/08474503
; Patent No. 5744446
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patencin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,503
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMU106CIP(3)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6500
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cdna sequence
US-08-474-503-2

Query Match 94.1%; Score 7234; DB 1; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
QY 1 ATRRYILGAVELSDWYMQSDLGELPVDARPPRPVKSPFNATSVVYKTLFVEFTVHLFN 60
DB 1 ATRRYILGAVELSDWYMQSDLGELPVDARPPRPVKSPFNATSVVYKTLFVEFTVHLFN 60
QY 61 IAKPRPPWMLGPTIOAEVYDVTVITLKNASHPVSLHAGVSYWKASGAEYDDQTSQ 120

DB 61 IAKPRPPWMLGPTIOAEVYDVTVITLKNASHPVSLHAGVSYWKASGAEYDDQTSQ 120
QY 121 REKEDDKVPPGGSHTYVQVQLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLICALVCR 180
DB 121 REKEDDKVPPGGSHTYVQVQLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLICALVCR 180
QY 181 EGSIAKEKQTLLHKFTLLFAVFDGKSWHSETTNSLMQDBDAASARAWPKMHTVNGVNR 240
DB 181 EGSIAKEKQTLLHKFTLLFAVFDGKSWHSETTNSLMQDBDAASARAWPKMHTVNGVNR 240
QY 241 SLPGILGCHRSYVHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITLTACTLL 300
DB 241 SLPGILGCHRSYVHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITLTACTLL 300
QY 301 MDLQGLLFCHISSHQHDGMEAYVKVDSCEPEZPOLRMKNNEEAEDYDDDLTDSMDVVR 360
DB 301 MDLQGLLFCHISSHQHDGMEAYVKVDSCEPEZPOLRMKNNEEAEDYDDDLTDSMDVVR 360
QY 361 DDNSPSFTQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPDDRYSKQYLANGFORIG 420
DB 361 DDNSPSFTQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPDDRYSKQYLANGFORIG 420
QY 421 RYKVKRFMAYTDTFTKTRAIQHESGILGPLLYGVGDTLLIIFKNQASRPNIYPHGI 480
DB 421 RYKVKRFMAYTDTFTKTRAIQHESGILGPLLYGVGDTLLIIFKNQASRPNIYPHGI 480
QY 481 TDVRPLYSRRLPKGVKHLXDFPILPGEIFKYKVTVTVEGPTKSDPRCLTRYSSFVYME 540
DB 481 TDVRPLYSRRLPKGVKHLXDFPILPGEIFKYKVTVTVEGPTKSDPRCLTRYSSFVYME 540
QY 541 RDLASGLIGPLLI CYKESVDQGNQIMSKDNVILFSVFDENRSWLTENIQRLPNPAG 600
DB 541 RDLASGLIGPLLI CYKESVDQGNQIMSKDNVILFSVFDENRSWLTENIQRLPNPAG 600
QY 601 VQLEDPEFOASNTMHSINGVVFDSLOLSVCLHVAVMYIISICAQTDFLSVFSGYTFKH 660
DB 601 VQLEDPEFOASNTMHSINGVVFDSLOLSVCLHVAVMYIISICAQTDFLSVFSGYTFKH 660
QY 661 KMVYEDTLTLFPFSGETVFMSENPGIWIIGCHNSDFRNRGMTALLKVSSCDKNTGYDE 720
DB 661 KMVYEDTLTLFPFSGETVFMSENPGIWIIGCHNSDFRNRGMTALLKVSSCDKNTGYDE 720
QY 721 DSYEDISAYLLSKNNAIEPRSF 742
DB 721 DSYEDISAYLLSKNNAIEPRSFQNGRHPSTROQFNATTIPENDIEKTDPMFAHRTMP 780
QY 743 742
DB 781 KIONVSSDLLMLLROSPTPHGLSLDLOEAKYETFDPSGAISSNNSLSEWTHFRPQ 840
QY 743 742
DB 841 LHHSGDMVFTPEESGLQLRLNEKLGTTAATLKKLDFKVSSTSNLSTIPSDNLAAGTDN 900
QY 743 742
DB 901 TSSLGPPSPVHYDSQDITLFGKKSPLTESGGPLSLSENNDSKILSEGLMNSQSSW 960
QY 743 742
DB 961 GKNVSTESGRLPFGKRAHGAPALLTKDNALFKVSIISLLTKNTKTSNNSATNRKTHIDGPSL 1020
QY 743 742
DB 1021 LIENSPVWQNILESDETEPKVTPLIHDRMLDKNATALLNHNMSKNTSSKNMVMVQK 1080
QY 743 742
DB 1081 KEGPIPPDAQNPMSFPKMLFLPESARWIQRTGKNSLNSGGQSPSKQLVSLGPEKSV 1140
QY 743 742
DB 1141 QNFLSEKNKVVVGKEFTKDVGLKEMVFPSSRNFLTLDNLNHNTHNQEKIQBIEK 1200

QY 743 ----- 742
Db 1201 KETLIQENVVLPOIHTVTGKTFKXNLFLLSTRQVSGSYEGAYAPVLQDFRSNDSTNR 1260
QY 743 ----- 742
Db 1261 TKKHTAHFKKBEENLEGLNQTQKIVEYACTTRISPNSTSQNFVTVORSKRALKQFRL 1320
QY 743 ----- 742
Db 1321 PLEBTELEKRIIVDDTSTQSKNKKHLTPSTLTQIDYNEKEKGAITQSPDCLTRSHSI 1380
QY 743 ----- 742
Db 1381 PQANRSPPIAKVSSFFSIRPIYILTRVLFDQNSSHLPAASVYRKXKDSGVQBSHFLQAKK 1440
QY 743 ----- 742
Db 1441 NNLSIALITLMTGQREVGSIGTSATNSVTYKKVENTVLPKDPKPTSGKVLLPKVHI 1500
QY 743 ----- 742
Db 1501 YKOLFPPTETSGSPGHLDLVEGSLLOQTEGAIKWNEANRPGKVPFLVATESSAKTPSK 1560
QY 743 ----- 742
Db 1561 LLDPLAWNHYGTQIPKEWKSQESPEKTAFFKKKDTILSLNACESNHAIAAINEGQNK 1620
QY 743 -----
Db 1621 EIEVTWAKQGRTERLCSQNPVVKRQREIRTRTTLQSDQEEIDYDDTISVEMKXEDFYI 1680
QY 743 -----
Db 787 DEDENQSPRSQKTRHYFIAAVERLWDYGMSSPHVLNRNQAQSGVQPKVVFQBEFTD 846
Db 1681 DEDENQSPRSQKTRHYFIAAVERLWDYGMSSPHVLNRNQAQSGVQPKVVFQBEFTD 1740
QY 847 GSFTQPLRGELNEHLGLLGPYIRAEVDNIWTFERNQASRPYSYSSLSIYEEQROGA 906
Db 1741 GSFTQPLRGELNEHLGLLGPYIRAEVDNIWTFERNQASRPYSYSSLSIYEEQROGA 1800
QY 907 EPRKNFVKNETKTYFWKVOHMAPTKQBFCKAWAYFSDVLEKDVHSGLLIGPLLVCHT 966
Db 1801 EPRKNFVKNETKTYFWKVOHMAPTKQBFCKAWAYFSDVLEKDVHSGLLIGPLLVCHT 1860
QY 967 NTLNPAHGRQVTVQBFALFFTFIDETKSWYFTENWERNCAPCNIQMEDPTFKENYRFA 1026
Db 1861 NTLNPAHGRQVTVQBFALFFTFIDETKSWYFTENWERNCAPCNIQMEDPTFKENYRFA 1920
QY 1027 INGYIMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIFSGHVTVTRKKEEYKXALYNLYP 1086
Db 1921 INGYIMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIFSGHVTVTRKKEEYKXALYNLYP 1980
QY 1087 GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGMSGHIRDQITAS 1146
Db 1981 GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGMSGHIRDQITAS 2040
QY 1147 GOYGQWAPKLARLHYSGSINASTKEPFSWKVLLAPMIHGIKTQGAROKFSSLIYSQ 1206
Db 2041 GOYGQWAPKLARLHYSGSINASTKEPFSWKVLLAPMIHGIKTQGAROKFSSLIYSQ 2100
QY 1207 FIIMYSLDGKCKQTYRGNSGTGLWVFGNVDSGKHNIFNPPIIARYIRLRLHPTHYSIRS 1266
Db 2101 FIIMYSLDGKCKQTYRGNSGTGLWVFGNVDSGKHNIFNPPIIARYIRLRLHPTHYSIRS 2160
QY 1267 TLRMELMGCDLNSCSPGLMGESKAIISDAQITASSYFTNPFATWSPSKARLHLQGRNANR 1326
Db 2161 TLRMELMGCDLNSCSPGLMGESKAIISDAQITASSYFTNPFATWSPSKARLHLQGRNANR 2220
QY 1327 PQVNNPKWILQVDFQKTMKVTGTTQGVKSLLTSMYKKEFLISSQDGHQWTLFFONGKV 1386
Db 2221 PQVNNPKWILQVDFQKTMKVTGTTQGVKSLLTSMYKKEFLISSQDGHQWTLFFONGKV 2280

QY 1387 KYVQGNQDSFTPVVNSLDPLLTRYLRIHQSWHVOIALRMEVLGCEAQDLY 1438
Db 2281 KYVQGNQDSFTPVVNSLDPLLTRYLRIHQSWHVOIALRMEVLGCEAQDLY 2332
RESULT 10
US-08-670-707A-2
; Sequence 2, Application US/08670707A
; Patent No. 5859204
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/670,707A
; FILING DATE: 26-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenlee, Lorraine L.
; REGISTRATION NUMBER: 27,894
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
; US-08-670-707A-2
Query Match 94.1%; Score 7234; DB 2; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
QY 1 ATRRYVLGAVELSWDMQSDLGELPVDARFPFRVPKSPFNTSVVYKKTLPVEFTVHLFN 60
Db 1 ATRRYVLGAVELSWDMQSDLGELPVDARFPFRVPKSPFNTSVVYKKTLPVEFTVHLFN 60
QY 61 IAKPRPPMGLLGPTIOAEVYDVTITLKNMASHVPSLHAGVSVYKASGAEYDDQTSQ 120
Db 61 IAKPRPPMGLLGPTIOAEVYDVTITLKNMASHVPSLHAGVSVYKASGAEYDDQTSQ 120
QY 121 REKEDDKVPFGSHYTVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180

Db	121	REXEDDKVFGGSHYVQVWVKENGPMASDPLCLTYSYLSHVLDVLDNSGLGALLVCR	180
Qy	181	EGSLAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKOHTVNGYVNR	240
Db	181	EGSLAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKOHTVNGYVNR	240
Qy	241	SLPGLIGCHRSKSVYVHWVGMGTTPVEHSIFLEGHTFLVRNHRQASLEISPIFTLTAQTL	300
Db	241	SLPGLIGCHRSKSVYVHWVGMGTTPVEHSIFLEGHTFLVRNHRQASLEISPIFTLTAQTL	300
Qy	301	MDLGOFLFCHISSHODGMEAYVYKVDSCPEEPOLRMKNNEEADYDDDLTDSMDVVF	360
Db	301	MDLGOFLFCHISSHODGMEAYVYKVDSCPEEPOLRMKNNEEADYDDDLTDSMDVVF	360
Qy	361	DDNSPSFQIRSVAKKHPTKWVHYIAAEEDMDYAPLVAPDDRSYKSOYLNNGPQIRIG	420
Db	361	DDNSPSFQIRSVAKKHPTKWVHYIAAEEDMDYAPLVAPDDRSYKSOYLNNGPQIRIG	420
Qy	421	RKYKVRFMAYTDTFTKTREAIQHSIGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHGI	480
Db	421	RKYKVRFMAYTDTFTKTREAIQHSIGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHGI	480
Qy	481	TDVRLYSRRLPKGVKHLKOPPLPGELFKYKWTVTVEDGPKSDPCLTRYSSFYVME	540
Db	481	TDVRLYSRRLPKGVKHLKOPPLPGELFKYKWTVTVEDGPKSDPCLTRYSSFYVME	540
Qy	541	RDLASGLIGPLLIICYESVDQRGNQIMSDKENVILFSVFDENRSWYLTENIQRLPNPAG	600
Db	541	RDLASGLIGPLLIICYESVDQRGNQIMSDKENVILFSVFDENRSWYLTENIQRLPNPAG	600
Qy	601	VQLEDPEFQASNMHSINGVYFDSLSQVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH	660
Db	601	VQLEDPEFQASNMHSINGVYFDSLSQVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH	660
Qy	661	KWYEDTLTLPPFSGEYFVMSMENPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDYIE	720
Db	661	KWYEDTLTLPPFSGEYFVMSMENPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDYIE	720
Qy	721	DSYEDISAYLLSKNNAIEPRSF-----	742
Db	721	DSYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTIPENDIEKTDWPAHRTMP	780
Qy	743	-----	742
Db	781	KIQNVSSDLLMLLRQSPTRPHGLSLSDIQEAKYETFFSDPSPGALDSNNSLSEMTHERPQ	840
Qy	743	-----	742
Db	841	LHSGDMVFTPESGIQLRLNEKLGTTAATELKLCLDFKVSSTSNLIISTIPSDNLAAGTDN	900
Qy	743	-----	742
Db	901	TSSIGPPSPMVHYDSQDSTTLFGKSSPLTESGGPLSUSEBNDKLLSEGLMNSQBSGW	960
Qy	743	-----	742
Db	961	GKNVSSSTESGRLFKGKRAHPALLTKONALFKVSI SLLTKNTKTSNNSATNRKTHDGPSL	1020
Qy	743	-----	742
Db	1021	LIENSPSVQNILESDBTEFKKVTPLIHDRMLMDKNATALRLNHSNKTTSKKNMERMVQK	1080
Qy	743	-----	742
Db	1081	KEGP1PPDAQNPDMSFFKMLFLPESARWIQRTGKNSLNSGGQSPKQVLSLGPBKSVEG	1140
Qy	743	-----	742
Db	1141	QNFLSEKKNVVGKGEFTKDVGLKEMVFPSSRNPLFTNLNLDLHNNHTHQEKKIQEIEK	1200
Qy	743	-----	742

1201	KETLIQENVVLPOIHTVTGTGNPMKNLFLLSLSTRQNVESYEGAYAPVLQDFRSLNDSTNR	1260
743	-----	742
1261	TKKHTAHFAKSGKEENLEGLGNQTKOIVEKIYACTTRISPNTSQONFVTOQRKALQKQRL	1320
743	-----	742
1321	PLBETELEKRIIVDDTSTQWKMKNHLPSTLTQIDYNEKEKGAIQSPSLDCLTRSHSI	1380
743	-----	742
1381	PQANRSPPLAKVSSPPSIRPIYLTRVLQDNSSHLPAASRYKKDKSGVBSHFLQGAKX	1440
743	-----	742
1441	NNLSAILTLEMTGDQREVGSLGTSATNSVYKKVENTVLKPLDKPTSGKVELLPKVIH	1500
743	-----	742
1501	YQKDLPTTSTNSGPHLDLVEGSLLOGTBGAIKWNEANRPKVPFLRVATESAKTPSK	1560
743	-----	742
1561	LLDPLAWNHYGTQIPKEWKSQEKSPKTAFFKKODTILSNACSNHATAINEQONKP	1620
743	-----SONPPVLKXHQRETRTRTLQSDQBEIDYDDDTISYEMKKEDFIY	786
1621	EIEVTWAKQRTBRLCSNPPVLKXHQRETRTRTLQSDQBEIDYDDDTISYEMKKEDFIY	1680
787	DEBQSPRSFOKTRHYFTAAVERLWDYGMSSSPHVLNRNRAOSGVSVPQKVVVFQFTD	846
1681	DEBQSPRSFOKTRHYFTAAVERLWDYGMSSSPHVLNRNRAOSGVSVPQKVVVFQFTD	1740
847	GSFTQPLYRSELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLYSEEDQOQA	906
1741	GSFTQPLYRSELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLYSEEDQOQA	1800
907	EPRKNFVKNETKTYFWKVOHMAPTKDEPDCKAWAYFSDVDLEKOVHSGLIGLPLVCHT	966
1801	EPRKNFVKNETKTYFWKVOHMAPTKDEPDCKAWAYFSDVDLEKOVHSGLIGLPLVCHT	1860
967	NTLNPAGROVTVQEFALFTTIDETKSWYFTENMERNCRAPCNIQMEDPTFKENYFPA	1026
1861	NTLNPAGROVTVQEFALFTTIDETKSWYFTENMERNCRAPCNIQMEDPTFKENYFPA	1920
1027	INGYIMDTLPLGVMAQOQIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKALYNLYP	1086
1921	INGYIMDTLPLGVMAQOQIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKALYNLYP	1980
1087	GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNCKQTPGLMGASHIRDFQITAS	1146
1981	GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNCKQTPGLMGASHIRDFQITAS	2040
1147	GOYQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQAGOKFSSLYISQ	1206
2041	GOYQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQAGOKFSSLYISQ	2100
1207	FIIMYSLDGKKQWYRGNSGTGLMVFFGNVDSGIGKINIFNPPIIARYIRLPHYSIRS	1266
2101	FIIMYSLDGKKQWYRGNSGTGLMVFFGNVDSGIGKINIFNPPIIARYIRLPHYSIRS	2160
1267	TLRMELMGCDLNSCMPGLMESKAI SDAQITASSYFTNMFTATWSPKARLHQGRSNWR	1326
2161	TLRMELMGCDLNSCMPGLMESKAI SDAQITASSYFTNMFTATWSPKARLHQGRSNWR	2220
1327	PQVNNPKWLOVDPKTKMKVTGVTQGVKSLTSMYKEFLISSQGHQWTLFPQNGKV	1386
2221	PQVNNPKWLOVDPKTKMKVTGVTQGVKSLTSMYKEFLISSQGHQWTLFPQNGKV	2280
1387	KVFGQNDSPFPVNVNSLDPLLTTRYLRTHPOSWHQIALRMEVLGCEAQDLY	1438
2281	KVFGQNDSPFPVNVNSLDPLLTTRYLRTHPOSWHQIALRMEVLGCEAQDLY	2332

Wed Dec 10 11:56:46 2003

RESULT 11

US-09-037-601-2
; Sequence 2, Application US/09037601
; Patent No. 6180371
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,601
; FILING DATE: 26-JUN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8089
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
; US-09-037-601-2

Query Match 94.1%; Score 7234; DB 3; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
Qy 1 ATRRYIYGAVELSDVMSDGLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 60
Db 1 ATRRYIYGAVELSDVMSDGLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 60
Qy 61 IAKRPPWMLLGGTIIQAEVYDVTIVITLKNMASHPVSLHVGVSYWKASGAEYDDQTSQ 120
Db 61 IAKRPPWMLLGGTIIQAEVYDVTIVITLKNMASHPVSLHVGVSYWKASGAEYDDQTSQ 120
Qy 121 REKEDDKVFGSGSHYVQVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR 180
Db 121 REKEDDKVFGSGSHYVQVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR 180

Qy 181 EGSIAKEKTOILHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVYNR 240
Db 181 EGSIAKEKTOILHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVYNR 240
Qy 241 SLPLIGCHRSVYVHVIGMGTTPVEHISIFLEGHTFLVRNHRQASLSIPITFLTAQLL 300
Db 241 SLPLIGCHRSVYVHVIGMGTTPVEHISIFLEGHTFLVRNHRQASLSIPITFLTAQLL 300
Qy 301 MDLQOFLFCHISSHQHDGMEAYVYKVDSCPEEPQLRMKNEEAEDYDDDLTDSMDVVRP 360
Db 301 MDLQOFLFCHISSHQHDGMEAYVYKVDSCPEEPQLRMKNEEAEDYDDDLTDSMDVVRP 360
Qy 361 DDNSPSFIQIRSVAKHKPTWVHYIAAEEDWDYAPLVLPADDDRSYKSOVLNNGPQRI 420
Db 361 DDNSPSFIQIRSVAKHKPTWVHYIAAEEDWDYAPLVLPADDDRSYKSOVLNNGPQRI 420
Qy 421 RYKVKRFMAYTDETFKTREAIQESGILGLPGLLYGEVGDITLLIIPKQASRPYNIYPHI 480
Db 421 RYKVKRFMAYTDETFKTREAIQESGILGLPGLLYGEVGDITLLIIPKQASRPYNIYPHI 480
Qy 481 TDVRLYSRRLPKGVKHLKOPPIIPGELFKYKWTVTVEDGPKSDPRLTRYYSFVNM 540
Db 481 TDVRLYSRRLPKGVKHLKOPPIIPGELFKYKWTVTVEDGPKSDPRLTRYYSFVNM 540
Qy 541 RDLASGLIGPLLI CYKESVDQGRNQIMSDKENVILFSVFDNRNRSWYLTENQRFNPNAG 600
Db 541 RDLASGLIGPLLI CYKESVDQGRNQIMSDKENVILFSVFDNRNRSWYLTENQRFNPNAG 600
Qy 601 VQLEDPEFQASNIHMSINGVYVDSIQLSVCLHEVAYWYILSIGAQDTFLSPFGSYTTPKH 660
Db 601 VQLEDPEFQASNIHMSINGVYVDSIQLSVCLHEVAYWYILSIGAQDTFLSPFGSYTTPKH 660
Qy 661 KMVEDTTLTLPFGSETVFMENPGLMILCHNSDFRNCGMTALLKYSSCDKNTGDYVE 720
Db 661 KMVEDTTLTLPFGSETVFMENPGLMILCHNSDFRNCGMTALLKYSSCDKNTGDYVE 720
Qy 721 DSYEDISAYLISKNNAIERPSF----- 742
Db 721 DSYEDISAYLISKNNAIERPSFQNSRHPSTKQKQFNATIPENDIEKTDPFARHTPMP 780
Qy 743 ----- 742
Db 781 KIQNVSSDLLMLLRQSPTPHGLSLSDIQAKEYETFSDDPSGAIIDNNLSLSEMTHERPQ 840
Qy 743 ----- 742
Db 841 LHSQDMVFTPEGLQLRLNEKLGTTAATELKKLDPKVSSSTNNLIITIPSDNLAAGTDN 900
Qy 743 ----- 742
Db 901 TSSLGPPSPVHYDSQDITTLFGKXSSPLTESGGPLSLSENNDSKLLSEGLMNSQESSW 960
Qy 743 ----- 742
Db 961 GKNVSTESGRLEPKGKRAHPALLTKDNALFKVISLTKNTKTSNNATNRKTHIDGPSL 1020
Qy 743 ----- 742
Db 1021 LIENSPSVQNILESDETFKVTPLIHDRMLADKNATALRLNHSNKTTSKKNMEMVQOK 1080
Qy 743 ----- 742
Db 1081 KEGPIPPDAQNPDMSFFKMLFLPESARWIQRTGKNSLNSGGQSPKQLVSLGPEKSVEG 1140
Qy 743 ----- 742
Db 1141 QNFLSEKNVVGKGEFTKDVGLKEMVFPSSRNLFLTNDLNLHNTHNQEKIQEIEK 1200
Qy 743 ----- 742
Db 1201 KETLIQENVVLPOIHTVTGTFKMFKNLFLLLSTRQNVESYEGAYAPVLQDFRSLNDSTNR 1260
Qy 743 ----- 742

Db 1261 TKKHTAHFSGKKGEENLEBGLNQTQKQIVEKIACTTRISNTSOQNFTORSKALKQFRL 1320
QY 743 ----- 742
Db 1321 PLEETELEKRIIVDDTSTQSKNMKHLTPSTLTQIDYNEKEGAIQTQSPDCLTRSHSI 1380
QY 743 ----- 742
Db 1381 PQANRSPLEIAKVSSPFSIRPIYLTVLFQDNSSHLPAASRYRKDSGVQESSHFLQAKK 1440
QY 743 ----- 742
Db 1441 NNLSLAILFLEMTGDQREVGSLGTSATNSVYKKVENTVLPKPDLPKTSKVELLPKVI 1500
QY 743 ----- 742
Db 1501 YQXDLPTETSNPGHLDLVEGSLLOCTGAIKWEANRPGKVPFLRVATESSAKTPSK 1560
QY 743 ----- 742
Db 1561 LLDPLAWNHYGTQIPKEEWSQESPEKTAFFKKDTILSLNACSNHAIANAINEQONKP 1620
QY 743 ----- 786
Db 1621 EIEVTWAKQGRBELCSQNPVLKRHQREITRTTLOSQDEEIDYDDTISVEMKKEPDFIY 1680
QY 787 DEBENOSPRSFQKTRHYFLAAVERLWDYQWSSPHVLRNRAQSGSVPOPKVVFQBFETD 846
Db 1681 DEBENOSPRSFQKTRHYFLAAVERLWDYQWSSPHVLRNRAQSGSVPOPKVVFQBFETD 1740
QY 847 GSFTQPLRGELNEHLLGPGYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYSEEDQOQA 906
Db 1741 GSFTQPLRGELNEHLLGPGYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYSEEDQOQA 1800
QY 907 EPRKNFVKNETKTYFWKVOHHMPTKDEFDCKAWAYFSDVLEKDVHSGILGPLLVCYT 966
Db 1801 EPRKNFVKNETKTYFWKVOHHMPTKDEFDCKAWAYFSDVLEKDVHSGILGPLLVCYT 1860
QY 967 NTLNPAHGRQVTVQEBALFTTIDETKSWYFTENMERNCRAPNIOMEDPTFKENTRFHA 1026
Db 1861 NTLNPAHGRQVTVQEBALFTTIDETKSWYFTENMERNCRAPNIOMEDPTFKENTRFHA 1920
QY 1027 INGYIMDTPLGLVMAQDQIRWYLLSGNSNENIHSIHFSGHVFTVRKKEEYKMAYNLYP 1086
Db 1921 INGYIMDTPLGLVMAQDQIRWYLLSGNSNENIHSIHFSGHVFTVRKKEEYKMAYNLYP 1980
QY 1087 GVPETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKCOTPLGNASGHIRDFOITAS 1146
Db 1981 GVPETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKCOTPLGNASGHIRDFOITAS 2040
QY 1147 GOYGOWAPKLARLHYSGINAWSTKEPFSWKVDLLAPMIIHGKTQGARQKTSLSYISQ 1206
Db 2041 GOYGOWAPKLARLHYSGINAWSTKEPFSWKVDLLAPMIIHGKTQGARQKTSLSYISQ 2100
QY 1207 FIIMYSIDGKKWQTYRGNSTGLTMVFFGNVDSSGIKENINFPPIIARYIRLPHYSIRS 1266
Db 2101 FIIMYSIDGKKWQTYRGNSTGLTMVFFGNVDSSGIKENINFPPIIARYIRLPHYSIRS 2160
QY 1267 TLRMELMGCDLNSCMPGLMESKAISDAQITASSYFTNMPATWSPSKARLHLQGRSNWR 1326
Db 2161 TLRMELMGCDLNSCMPGLMESKAISDAQITASSYFTNMPATWSPSKARLHLQGRSNWR 2220
QY 1327 PQVNNPKMLQVDPQKTMKVTGVTQGVKSLTSMYKFEFLISSQDGHQWTLFPQNGKY 1386
Db 2221 PQVNNPKMLQVDPQKTMKVTGVTQGVKSLTSMYKFEFLISSQDGHQWTLFPQNGKY 2280
QY 1387 KVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWHQIALRMEVLGCEAQDLY 1438
Db 2281 KVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWHQIALRMEVLGCEAQDLY 2332

US-09-315-179-2
; Sequence 2, Application US/09315179
; Patent No. 6376463
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95H
; CURRENT APPLICATION NUMBER: US/09/315,179
; CURRENT FILING DATE: 1999-05-20
; EARLIER APPLICATION NUMBER: U.S. 09/037,601
; EARLIER FILING DATE: 1998-03-10
; EARLIER APPLICATION NUMBER: U.S. 08/670,707
; EARLIER FILING DATE: 1996-06-26
; EARLIER APPLICATION NUMBER: PCT/US97/111155
; EARLIER FILING DATE: 1997-06-26
; EARLIER APPLICATION NUMBER: PCT/US94/13200
; EARLIER FILING DATE: 1994-11-15
; EARLIER APPLICATION NUMBER: U.S. 08/212,133
; EARLIER FILING DATE: 1994-03-11
; EARLIER APPLICATION NUMBER: U.S. 07/864,004
; EARLIER FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-179-2

Query Match 94.1%; Score 7234; DB 4; Length 2332;

Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
QY 1 ATRYYLGAVELSWDMQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 60
DB 1 ATRYYLGAVELSWDMQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 60
QY 61 IAKPRPPNMLGPTTQAEVYDVTVTITLKNMASHPVSLHAVGVSYWKAEGASYDDQTSQ 120
DB 61 IAKPRPPNMLGPTTQAEVYDVTVTITLKNMASHPVSLHAVGVSYWKAEGASYDDQTSQ 120
QY 121 REKEDKVPFGSSHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
DB 121 REKEDKVPFGSSHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
QY 181 EGSIAKEKTQTLHKFTILLFAVDFDEGKSWSETKNSLMQDRDAASARAWPQHTVGVYNR 240
DB 181 EGSIAKEKTQTLHKFTILLFAVDFDEGKSWSETKNSLMQDRDAASARAWPQHTVGVYNR 240
QY 241 SLPLGIGCHRSYVYMHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPTITLTAQTL 300
DB 241 SLPLGIGCHRSYVYMHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPTITLTAQTL 300
QY 301 MDLQGFLLCHTSSHQHDMGEAYVKYDSCPEEPQLEKNEEAEYDDDLTDSEMDVVRP 360
DB 301 MDLQGFLLCHTSSHQHDMGEAYVKYDSCPEEPQLEKNEEAEYDDDLTDSEMDVVRP 360
QY 361 DDNSPSFTQIRSVAKKHPTWVHYIAAEEEDMDYAPLVLPDDRYSKSYOLNNGPQIRIG 420
DB 361 DDNSPSFTQIRSVAKKHPTWVHYIAAEEEDMDYAPLVLPDDRYSKSYOLNNGPQIRIG 420
QY 421 RYKTKVRFMAYTDETFKTRERAIQHESGILGPLYGEVGDTLIIIPKQASRPNIYPHGI 480
DB 421 RYKTKVRFMAYTDETFKTRERAIQHESGILGPLYGEVGDTLIIIPKQASRPNIYPHGI 480
QY 481 TDVRPLYSRRLPKGVKHLKDPFLLPQRIKTKVTVVEDEGPTKSDPRCLTRYYSFVWME 540
DB 481 TDVRPLYSRRLPKGVKHLKDPFLLPQRIKTKVTVVEDEGPTKSDPRCLTRYYSFVWME 540
QY 541 RDLASGLIGPLLIICYKESVDQRCNOJMSDKRNVILSFVDENRSWLTENIORPLNPAG 600
DB 541 RDLASGLIGPLLIICYKESVDQRCNOJMSDKRNVILSFVDENRSWLTENIORPLNPAG 600

QY 601 VQLEDPFQASNMHSGVDFSLQSVCLHEVAYWYILSIGAQDTFLSVFSGYTFKH 660
DB 601 VQLEDPFQASNMHSGVDFSLQSVCLHEVAYWYILSIGAQDTFLSVFSGYTFKH 660
QY 661 KMYVEDTLTLPFPGSETVFMSENPGLWILCHNSDFRNGMTALLKVSSCDKNTGDIYE 720
DB 661 KMYVEDTLTLPFPGSETVFMSENPGLWILCHNSDFRNGMTALLKVSSCDKNTGDIYE 720
QY 721 DSYEDISAYLLSKNAIEPRSP----- 742
DB 721 DSYEDISAYLLSKNAIEPRSPQNSRHPSTRQKQFNATIPENDIEKTDWFAHRTMP 780
QY 743 ----- 742
DB 781 KIQNVSSDLLMLLRQSPHPGLSLDIQEAKEYTFSDPSPGAIDNSNLSLSEMTFRPQ 840
QY 743 ----- 742
DB 841 LHSQDMVFTPESGIQLRLNEKLGTAATELKLDPKVSSTNNLISTIPSDNLAAGTDN 900
QY 743 ----- 742
DB 901 TSSLGPPSPMHHYDSQDITTLFGKXSPHTBSGGPLSLSEENDSKLLSEGLMNSQESSM 960
QY 743 ----- 742
DB 961 GKNVSTESGRLFKGRAHPALLTKDNALFKVISLTKTKNTKTSNNSATNRKTHIDGPSL 1020
QY 743 ----- 742
DB 1021 LIENSPVWONILESDTEBFKVTPLIHRMLMDKNATALRLNHNKNTTSKQNMENVOOK 1080
QY 743 ----- 742
DB 1081 KEGPIPPDAQNPMDFKMLFLPESARWIORTHGKNSLNGOGSPKQLVSLGPEKSVBG 1140
QY 743 ----- 742
DB 1141 QNFLSEKNKVVVGKBEFTKDVGLKEMVPPSRNLFTNLNLDLHNNHNTKOEIEBK 1200
QY 743 ----- 742
DB 1201 KETLIQENVLPQIHTVTGTFKFMKMLFLLSTRQNVGSEYEGAYAPVLQDFRSLNDSTNR 1260
QY 743 ----- 742
DB 1261 TKKHTAHFSKKEENLEGLGNQTKQIYEKACTTIPSTSQNFVQTSKRALQKPL 1320
QY 743 ----- 742
DB 1321 PLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKEGAIQSPSLDCLTRSHSI 1380
QY 743 ----- 742
DB 1381 PQANRSPPLIAKVSSPFSIRPIYLRVLFQDNSSHLPAASRYKSDSGVQESSHFLQAKK 1440
QY 743 ----- 742
DB 1441 NNLSLAILLEMTGDQREVGLSGTSATNSVYKKVENTVLPKPDLPKTSKVELLPKVIH 1500
QY 743 ----- 742
DB 1501 YQXDLFTPTNSGPHGLDLVEGSLLOTEGAIKWNEANRPGKVPFLRVATESAKTPSK 1560
QY 743 ----- 742
DB 1561 LLDPLADNHYGTQIPKEWKSQESPEKTAFAKKDTILSLNACESNHAIAANEQONKP 1620
QY 743 -----SONPVLKRHOREITRITLQSDQEEIDYDDTISVEMKEDPDYI 786
DB 1621 EIEVTWAKQRTKLCQNPVLRHOREITRITLQSDQEEIDYDDTISVEMKEDPDYI 1680
QY 787 DEDENQSPRSFQKTRHYFAAVERLWDYGMSSSPHVLNRNAQSGSVFPQKVVQBFETD 846

DB 1681 DEDENQSPRSFQKTRHYFAAVERLWDYGMSSSPHVLNRNAQSGSVFPQKVVQBFETD 1740
QY 847 GSTQPLXRGELMEHLLGLGPYIRAEVEDNIMVTFRNOASRPYSFSSLSIYSEDQOGA 906
DB 1741 GSTQPLXRGELMEHLLGLGPYIRAEVEDNIMVTFRNOASRPYSFSSLSIYSEDQOGA 1800
QY 907 EPRKNFVKNETTYFWKQOHMAPTKDEPDCAKAWAFSDVDLEKDVHSLIGPLVCHT 966
DB 1801 EPRKNFVKNETTYFWKQOHMAPTKDEPDCAKAWAFSDVDLEKDVHSLIGPLVCHT 1860
QY 967 NTLNPAHGRQVTVQEPALFTTIDETKSWYFTENMEENCRAPCNIOEMEDTTFKENTRPHA 1026
DB 1861 NTLNPAHGRQVTVQEPALFTTIDETKSWYFTENMEENCRAPCNIOEMEDTTFKENTRPHA 1920
QY 1027 INGYIMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEBKALYNLYP 1086
DB 1921 INGYIMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEBKALYNLYP 1980
QY 1087 GVPETVEMLPKAGIWEVECLIGBHLHAGMSTLFLVYSNKCOTPLGMASGHIRDFQITAS 1146
DB 1981 GVPETVEMLPKAGIWEVECLIGBHLHAGMSTLFLVYSNKCOTPLGMASGHIRDFQITAS 2040
QY 1147 GOYQWAPKLARLHYSGSINAWSTKBPFSWIKVDLLAPMIHGIKTQCARQKFSLSYISQ 1206
DB 2041 GOYQWAPKLARLHYSGSINAWSTKBPFSWIKVDLLAPMIHGIKTQCARQKFSLSYISQ 2100
QY 1207 FIIMYSLDGKKWQYRGNSTGTLMVFFGNVDSSGKINI FNPPIIARIYIRLHPHYSIRS 1266
DB 2101 FIIMYSLDGKKWQYRGNSTGTLMVFFGNVDSSGKINI FNPPIIARIYIRLHPHYSIRS 2160
QY 1267 TLRMELMGCDLNSCMPLGESKALSDAQITASSYFTNMATWSPSKARLHLOQESNWR 1326
DB 2161 TLRMELMGCDLNSCMPLGESKALSDAQITASSYFTNMATWSPSKARLHLOQESNWR 2220
QY 1327 PQVNNPKEWLQVDFQKTMKVTVGTTQGVKSLTSMYKFEFLISSQDGHQWTLFFQNGKV 1386
DB 2221 PQVNNPKEWLQVDFQKTMKVTVGTTQGVKSLTSMYKFEFLISSQDGHQWTLFFQNGKV 2280
QY 1387 KVFQNGQDSFTPVVNSLDPPLLTRYLRIHPOSWHQIALRMEVLGCEAODLY 1438
DB 2281 KVFQNGQDSFTPVVNSLDPPLLTRYLRIHPOSWHQIALRMEVLGCEAODLY 2332

RESULT 13
US-09-523-656-2
; Sequence 2, Application US/09523656
; Patent No. 6458563
; GENERAL INFORMATION:
; APPLICANT: Lollar S., John
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-951
; CURRENT APPLICATION NUMBER: US/09/523,656
; CURRENT FILING DATE: 2000-03-10
; EARLIER APPLICATION NUMBER: 09/037,601
; EARLIER FILING DATE: 1998-03-10
; EARLIER APPLICATION NUMBER: 08/670,707
; EARLIER FILING DATE: 1996-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-523-656-2

Query Match 94.1%; Score 7234; DB 4; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
QY 1 ATRRYLGAVELSWDYMQSDLGELPVDARPPRPVPSFPNTSVVYKTLFVEFTVHLEN 60
DB 1 ATRRYLGAVELSWDYMQSDLGELPVDARPPRPVPSFPNTSVVYKTLFVEFTVHLEN 60

Db 2221 PQVNNPEKLVDFQKTMKVTGVTQGVKSLLTSMYKFLISSQDGHQWTLFFQNGKY 2280
Qy 1387 KVFQGNQDSFTPVVNSLDPLLTRYLRIHQSWHQIALRMEVLGCEAODLY 1438
Db 2281 KVFQGNQDSFTPVVNSLDPLLTRYLRIHQSWHQIALRMEVLGCEAODLY 2332

RESULT 14
PCT-US93-03275-4
; Sequence 4, Application PC/TUS9303275
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Runge, Marshall S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03275
; FILING DATE: 19930407
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU 106PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cdna sequence
PCT-US93-03275-4

Query Match . 94.1%; Score 7234; DB 5; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
Qy 1 ATRRYLGAVELSWDYMQSDGLPVDAREPPRPVKSPFNTSVVYKTKTLFVEFTVHLFN 60
Db 1 ATRRYLGAVELSWDYMQSDGLPVDAREPPRPVKSPFNTSVVYKTKTLFVEFTVHLFN 60
Qy 61 IAKPRPPWMLGFTTQAEVYDVTWITLKNMASHPVSLHAGVSVYKASGAYDDQTSQ 120
Db 61 IAKPRPPWMLGFTTQAEVYDVTWITLKNMASHPVSLHAGVSVYKASGAYDDQTSQ 120
Qy 121 REKEDDKVFGGSHYVWQVLKNGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR 180
Db 121 REKEDDKVFGGSHYVWQVLKNGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR 180

Qy 181 EGS LAKEKCTQTLHKFTLLFAVDEBGSWHSEKSNLMDORDAASARAWPKMHTVNGYVNR 240
Db 181 EGS LAKEKCTQTLHKFTLLFAVDEBGSWHSEKSNLMDORDAASARAWPKMHTVNGYVNR 240
Qy 241 SLPGILGCHRKSVYHVIQMGTTTPEVHSIFLEGHTEFLVRNHRQASLEISPTITLTAQTL 300
Db 241 SLPGILGCHRKSVYHVIQMGTTTPEVHSIFLEGHTEFLVRNHRQASLEISPTITLTAQTL 300
Qy 301 MDLGGFLFCHISSHODGMEAYVKVDSCEEPOLRMKNNEAEEDDDDLTDSMDVVRP 360
Db 301 MDLGGFLFCHISSHODGMEAYVKVDSCEEPOLRMKNNEAEEDDDDLTDSMDVVRP 360
Qy 361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLPDDRYSKYSLNNGPQIRIG 420
Db 361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLPDDRYSKYSLNNGPQIRIG 420
Qy 421 RYKVKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGTLLIFKNQASRPNIYPHGI 480
Db 421 RYKVKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGTLLIFKNQASRPNIYPHGI 480
Qy 481 TDVRPLYSRRLPKGVKHLKDFPILPGEIFPKYKVTVTVEGPTKSDPRCLTRYSSFFVME 540
Db 481 TDVRPLYSRRLPKGVKHLKDFPILPGEIFPKYKVTVTVEGPTKSDPRCLTRYSSFFVME 540
Qy 541 RDLASGLIGPLLICYKESVDORGNQIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 600
Db 541 RDLASGLIGPLLICYKESVDORGNQIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 600
Qy 601 VQLEDPEFQASNMHSINGVFDLSQLSVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 660
Db 601 VQLEDPEFQASNMHSINGVFDLSQLSVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKH 660
Qy 661 KMVYEDTLTLFPFSGETVFMENPGLWILGCHNSDFRNGMTALLKVSSCDKNTGYE 720
Db 661 KMVYEDTLTLFPFSGETVFMENPGLWILGCHNSDFRNGMTALLKVSSCDKNTGYE 720
Qy 721 DSYEDISAYLLSKNAIEPRSF----- 742
Db 721 DSYEDISAYLLSKNAIEPRSFQNSRHPSTROKFNATTIPENDIEKTDPAHRTMP 780
Qy 743 ----- 742
Db 781 KIQNVSSDLLMLLRQSPTHGLSLSDIQAKEYTSDDPSPGAIDSNNSLSETHFRPQ 840
Qy 743 ----- 742
Db 841 LHSGLMVFTPSGLQLRLNEKLGTATATLKLDFKVSSTSNLLISTIPSDNLAAGTDN 900
Qy 743 ----- 742
Db 901 TSSLGPPSPMPVHYDSQLDTTLFGKKSPLTESGGPLSISENNDSKLLSGLMNSQSSW 960
Qy 743 ----- 742
Db 961 GKNVSSSTESGRLFKGRAHPALLFKDNALPKVISLKLTKNTKNNATNKRTHIDGFSL 1020
Qy 743 ----- 742
Db 1021 LIENSPSVWQNILESDETFKVTPLIHDRMLMDKNATRLNHNMGKTTSSKNMEMVQOK 1080
Qy 743 ----- 742
Db 1081 KEGPIPPDAQNPDMSFFKMLFLPESARWIQRTGKNLSNGQGPSKQLVSLGPEKSVEG 1140
Qy 743 ----- 742
Db 1141 QNFLSEKNKVYVGKGFVKDVGKEMVFPSSRNFLTNLNLHNHNTNHEKKIQEIEK 1200
Qy 743 ----- 742
Db 1201 KETLIQENVVLPIHTVTGTKNFMKNLFLSTRQNVGSEYAGYAPVQLDFRSLNDSTR 1260
Qy 743 ----- 742

```

Db      1261  TKKHTAHFKGGEENLEGLGNQTKQIVKEYACTTRISPNTSQONFVQSRKALQKFL 1320
QY      743  ----- 742
Db      1321  PLEETELEKRIIVDDTSTQWKNMKHLTPSTLTQIDYNEKEGAIQTQSLDCLTSHSI 1380
QY      743  ----- 742
Db      1381  PQANRSPPLIAKVSSPFSIRPIVLRVLFQDNSSHLPAASRYRKDQSGVQESSHFLQAKK 1440
QY      743  ----- 742
Db      1441  NNLSLAILTEMTGDQREVGSGLTSATNSVTYKKVENTVLPKPDLPKTSKGVKELLPKVHI 1500
QY      743  ----- 742
Db      1501  YQKDLFPTETSNQSPGHLDLVEGSLIQCTEGAIKNEANRPKGVPLRVATESSAKTPSK 1560
QY      743  ----- 742
Db      1561  LLDPLANDNHYGTQIPKEBWKQSKSPEKTAFFKKDTILSLNACESNHAIAINEQNKP 1620
QY      743  ----- 742
Db      1621  EIEVTWAKQGRTERLCSQNPVLRKHOREITRRTLOSDEEIDYDDTISVEMKKEFDIY 1680
QY      743  ----- 742
Db      1681  DEENQSPRSFQKTRHYFIAAVERLDYGMSSSPHLNRAGSGVPOPKKVVFQEFDT 846
QY      743  ----- 742
Db      1681  DEENQSPRSFQKTRHYFIAAVERLDYGMSSSPHLNRAGSGVPOPKKVVFQEFDT 1740
QY      743  ----- 742
Db      1741  GSFTQPLRGELNEHLGLGPIYRAVEDNIMVTFRNQASRPYSFYSSLSIYSEEDORQGA 906
QY      743  ----- 742
Db      1741  GSFTQPLRGELNEHLGLGPIYRAVEDNIMVTFRNQASRPYSFYSSLSIYSEEDORQGA 1800
QY      743  ----- 742
Db      1801  EPRKNFVKNETKTYFWKQVHMAPTKDFDCKAWAYFSDVLEKDVHSGSLIGPLVCHT 966
QY      743  ----- 742
Db      1801  EPRKNFVKNETKTYFWKQVHMAPTKDFDCKAWAYFSDVLEKDVHSGSLIGPLVCHT 1860
QY      743  ----- 742
Db      1861  NTLPNAGROVTVQEFALFTTIDETKSWYFTENMERNCAPCNIOEMEDFTFKENTRFA 1026
QY      743  ----- 742
Db      1861  NTLPNAGROVTVQEFALFTTIDETKSWYFTENMERNCAPCNIOEMEDFTFKENTRFA 1920
QY      743  ----- 742
Db      1861  NTLPNAGROVTVQEFALFTTIDETKSWYFTENMERNCAPCNIOEMEDFTFKENTRFA 1980
QY      743  ----- 742
Db      1861  NTLPNAGROVTVQEFALFTTIDETKSWYFTENMERNCAPCNIOEMEDFTFKENTRFA 2040
QY      743  ----- 742
Db      1861  NTLPNAGROVTVQEFALFTTIDETKSWYFTENMERNCAPCNIOEMEDFTFKENTRFA 2100
QY      743  ----- 742
Db      1861  NTLPNAGROVTVQEFALFTTIDETKSWYFTENMERNCAPCNIOEMEDFTFKENTRFA 2160
QY      743  ----- 742
Db      1861  NTLPNAGROVTVQEFALFTTIDETKSWYFTENMERNCAPCNIOEMEDFTFKENTRFA 2220
QY      743  ----- 742
Db      1861  NTLPNAGROVTVQEFALFTTIDETKSWYFTENMERNCAPCNIOEMEDFTFKENTRFA 2280
QY      743  ----- 742
Db      1861  NTLPNAGROVTVQEFALFTTIDETKSWYFTENMERNCAPCNIOEMEDFTFKENTRFA 2340
QY      743  ----- 742

```

RESULT 15

```

PCT-US94-13200-2
; Sequence 2, Application PC/TUS9413200
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13200
; FILING DATE: 15-NOV-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Babst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMD106CIP(2)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
;
PCT-US94-13200-2

```

```

Query Match      94.1%; Score 7234; DB 5; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;

QY      1  ATRRYLGAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTVHLFN 60
Db      1  ATRRYLGAVELSWDMQSDLGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTVHLFN 60
QY      61  IAKPRPMMGLLPTTQAEVYDVTVTLLKNMASHPVSLHAGVSYKASGAEYDDQTSQ 120
Db      61  IAKPRPMMGLLPTTQAEVYDVTVTLLKNMASHPVSLHAGVSYKASGAEYDDQTSQ 120
QY      121  REKEDDKVPFGGSHTYVWQVLEKNGPMASDPLCLTYSYLSHVDLVKDLASGLIGALLVCR 180
Db      121  REKEDDKVPFGGSHTYVWQVLEKNGPMASDPLCLTYSYLSHVDLVKDLASGLIGALLVCR 180
QY      181  EGS LAKEKQTLLHKFTILLFAVDEGKSWHSETKNSLMQDRDAASARAWPQHTVNGVNR 240
Db      181  EGS LAKEKQTLLHKFTILLFAVDEGKSWHSETKNSLMQDRDAASARAWPQHTVNGVNR 240
QY      241  SUPGLIGCHRSYVWHVIGMTTPEVHSIFLGHTFLVNRHRCASLEISPIITELTAOTLL 300
Db      241  SUPGLIGCHRSYVWHVIGMTTPEVHSIFLGHTFLVNRHRCASLEISPIITELTAOTLL 300
QY      301  MDLQGLFLLCHISSHQDGMKAYKVDSCPEEPQLMKNNHAEYDDDLTDSMDVVR 360
Db      301  MDLQGLFLLCHISSHQDGMKAYKVDSCPEEPQLMKNNHAEYDDDLTDSMDVVR 360
QY      361  DDNNSPFSFIQIRSVAKKHPTWVHYTAAEEDWDYAPLVLPDDRYSKQYLNNGPQIRG 420

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 16:55:27 ; Search time 43 Seconds
(without alignments)
6219.640 Million cell updates/sec

Title: US-10-006-091-1
Perfect score: 7691
Sequence: 1 ATRRYLGAVELSWDYMOSD.....VHQIALRMEVLCGEAQDLY 1438

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	7691	100.0	1438	14	US-10-006-091-1
2	7691	100.0	1438	14	Sequence 1, Appli
3	7691	100.0	1438	15	Sequence 1, Appli
4	7674	99.8	1471	14	US-10-095-718-2
5	7234	94.1	2332	10	US-09-957-641-2
6	7234	94.1	2332	12	US-10-131-510A-2
7	7234	94.1	2332	12	US-10-187-319-2
8	7227	94.0	2351	12	US-10-133-907-4
9	7227	94.0	2351	15	US-10-132-829-4
10	7227	94.0	2351	15	US-10-172-712-27
11	6599	85.8	1431	14	US-10-095-718-4
12	6489	84.4	1443	12	US-10-131-510A-39
13	6489	84.4	1443	15	US-10-187-319-39
14	6227	81.0	2133	12	US-10-131-510A-37
15	6227	81.0	2133	15	US-10-187-319-37

16	6192	80.5	2319	12	US-10-131-510A-6
17	6192	80.5	2319	15	US-10-187-319-6
18	2389.5	31.1	2224	15	US-10-115-563-14
19	2389.5	31.1	2224	15	US-10-172-712-31
20	1742.5	22.7	1160	12	US-10-137-870-234
21	1742.5	22.7	1160	12	US-10-140-018-234
22	1742.5	22.7	1160	12	US-10-140-021-234
23	1742.5	22.7	1160	12	US-10-140-274-234
24	1742.5	22.7	1160	12	US-10-140-471-234
25	1742.5	22.7	1160	12	US-10-140-807-234
26	1742.5	22.7	1160	12	US-10-140-922-234
27	1742.5	22.7	1160	12	US-10-140-924-234
28	1742.5	22.7	1160	12	US-10-140-926-234
29	1742.5	22.7	1160	12	US-10-141-698-234
30	1742.5	22.7	1160	12	US-10-141-702-234
31	1742.5	22.7	1160	12	US-10-141-704-234
32	1742.5	22.7	1160	12	US-10-142-421-234
33	1742.5	22.7	1160	12	US-10-142-432-234
34	1742.5	22.7	1160	12	US-10-142-767-234
35	1742.5	22.7	1160	12	US-10-143-033-234
36	1742.5	22.7	1160	12	US-10-144-994-234
37	1742.5	22.7	1160	12	US-10-145-628-234
38	1742.5	22.7	1160	12	US-10-145-631-234
39	1742.5	22.7	1160	12	US-10-145-633-234
40	1742.5	22.7	1160	12	US-10-145-746-234
41	1742.5	22.7	1160	12	US-10-145-748-234
42	1742.5	22.7	1160	12	US-10-145-823-234
43	1742.5	22.7	1160	12	US-10-145-826-234
44	1742.5	22.7	1160	12	US-10-145-870-234
45	1742.5	22.7	1160	12	US-10-145-876-234

ALIGNMENTS

RESULT 1
US-10-006-091-1
; Sequence 1, Application US/10006091
; Publication No. US20020102730A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255.1
; CURRENT APPLICATION NUMBER: US/10/006,091
; CURRENT FILING DATE: 2001-12-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-006-091-1

Query Match	100.0%;	Score 7691;	DB 14;	Length 1438;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1438;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATRRYLGAVELSWDYMOSDLGELPVDARPPRPVKSPFNTSVYKTLFVFTVHLFN	60	
Db	1	ATRRYLGAVELSWDYMOSDLGELPVDARPPRPVKSPFNTSVYKTLFVFTVHLFN	60	
Qy	61	IAKPRPFWMLGFTTQAEYVDTVTWITLKNMASHPSLHVGVSYMKASGAEYDDQTSQ	120	
Db	61	IAKPRPFWMLGFTTQAEYVDTVTWITLKNMASHPSLHVGVSYMKASGAEYDDQTSQ	120	
Qy	121	REKEDDKVFGGSHYVQVLYKENGPMASDPLCLTYSYLSHVDLVKDLNGLGALLVCR	180	

Db 121 REKEDKVPFGSGSHYVWQVLKENGPMASDPLCLTYSYLSHVLDLKDLSGLIGALLVCR 180
Qy 181 EGSIAKEKTTQTLHKETILLFAVDECKSWHSETKNSLMQDRDAASARAPKMTVNGYNR 240
Db 181 EGSIAKEKTTQTLHKETILLFAVDECKSWHSETKNSLMQDRDAASARAPKMTVNGYNR 240
Qy 241 SLPLGLIGCHRSVYWHVIGMGTTPPEVHSIFLFGHTFLVRNHRQASLEISPIFTLTAQTLL 300
Db 241 SLPLGLIGCHRSVYWHVIGMGTTPPEVHSIFLFGHTFLVRNHRQASLEISPIFTLTAQTLL 300
Qy 301 MDLGGFLFLFCHISSHQHDGMEAYVKVDCSPESPQLRMKNNEAEYDDDLTDSMDVVRP 360
Db 301 MDLGGFLFLFCHISSHQHDGMEAYVKVDCSPESPQLRMKNNEAEYDDDLTDSMDVVRP 360
Qy 361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLPDDRYSKYQYLANGPQIG 420
Db 361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLPDDRYSKYQYLANGPQIG 420
Qy 421 RYKVKVRFWAYTDETFKTRAIQHESGILGPLLYGEVGDITLLIIFKNQASRPYNIYPHGI 480
Db 421 RYKVKVRFWAYTDETFKTRAIQHESGILGPLLYGEVGDITLLIIFKNQASRPYNIYPHGI 480
Qy 481 TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKMTVTVEDGPKSDPRCLTRYISSFVNME 540
Db 481 TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKMTVTVEDGPKSDPRCLTRYISSFVNME 540
Qy 541 RLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIORELPNPAG 600
Db 541 RLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIORELPNPAG 600
Qy 601 VOLEDPEFOASNIMHSINGVYVDFSLQSVCLHEVAYWYILSICAQTDPLSFVSGYTFKH 660
Db 601 VOLEDPEFOASNIMHSINGVYVDFSLQSVCLHEVAYWYILSICAQTDPLSFVSGYTFKH 660
Qy 661 KMVEDTLTLFPFSGTVMFMSMENPGLNWLIGCHNSDFRNGMTALLKVSCKNTQDYTE 720
Db 661 KMVEDTLTLFPFSGTVMFMSMENPGLNWLIGCHNSDFRNGMTALLKVSCKNTQDYTE 720
Qy 721 DSYEDISAYLLSKNNAIEPRSSONPPVLKREHREITRTTLOSQDSEIYDDTISVEMKK 780
Db 721 DSYEDISAYLLSKNNAIEPRSSONPPVLKREHREITRTTLOSQDSEIYDDTISVEMKK 780
Qy 781 EDFDIYDEBENQSPRFQKTRHYFLAAVERLMDYGMSSPHVLRNRAQSGSVQPKKV 840
Db 781 EDFDIYDEBENQSPRFQKTRHYFLAAVERLMDYGMSSPHVLRNRAQSGSVQPKKV 840
Qy 841 FQFTDGSFTQPIYRGELNEHLGLGPYIRAEVEDNIMVTFNRQAARPYSFYSSLSIYEE 900
Db 841 FQFTDGSFTQPIYRGELNEHLGLGPYIRAEVEDNIMVTFNRQAARPYSFYSSLSIYEE 900
Qy 901 DQOQGAEPKKNFKPNETKTYFWKVOHMAPTKDEEDCKAWAYPSDVLKDVHSGLIGP 960
Db 901 DQOQGAEPKKNFKPNETKTYFWKVOHMAPTKDEEDCKAWAYPSDVLKDVHSGLIGP 960
Qy 961 LLVCHTNTLMPAHCROVTVQEPALFTTIFDETBSWYFTENNERNCRAPCNIQMEDTFKE 1020
Db 961 LLVCHTNTLMPAHCROVTVQEPALFTTIFDETBSWYFTENNERNCRAPCNIQMEDTFKE 1020
Qy 1021 NYRFHAINGYIMDTLPLGVMAQDQIRWYLLSGNSNENIHSIFSHGHVFTVRKKEBKMA 1080
Db 1021 NYRFHAINGYIMDTLPLGVMAQDQIRWYLLSGNSNENIHSIFSHGHVFTVRKKEBKMA 1080
Qy 1081 LYNLYPGVFTVEMLPKAGIWRVECLIGEHLAGMSTLFLVSNKQOTPLGMASGHIRD 1140
Db 1081 LYNLYPGVFTVEMLPKAGIWRVECLIGEHLAGMSTLFLVSNKQOTPLGMASGHIRD 1140
Qy 1141 FOITASQYQWAPKARLHYSGSINAWSTKEPFWMLKVDLLAPMIITHGKTQAGOKFS 1200
Db 1141 FOITASQYQWAPKARLHYSGSINAWSTKEPFWMLKVDLLAPMIITHGKTQAGOKFS 1200
Qy 1201 SLXISQFIIMYSLDGKKWQYTRGNSTGTLVFFGNVDSSGIKHNI FNPPIIARIYLRHPT 1260
Db 1201 SLXISQFIIMYSLDGKKWQYTRGNSTGTLVFFGNVDSSGIKHNI FNPPIIARIYLRHPT 1260

Qy 1261 HYSIRSTLRMELMGCDLNSCSPGLMESKAISDAQITASSYFTNNFATWSPSKARLHQ 1320
Db 1261 HYSIRSTLRMELMGCDLNSCSPGLMESKAISDAQITASSYFTNNFATWSPSKARLHQ 1320
Qy 1321 RSNARPOVNNPKWLQVDFOKTKMKVTGTTQGVKSLTSMYKVEFLISSQDGHQWTLF 1380
Db 1321 RSNARPOVNNPKWLQVDFOKTKMKVTGTTQGVKSLTSMYKVEFLISSQDGHQWTLF 1380
Qy 1381 FONGKVKVFCQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWHQAIALRMEVLGCEAQDLY 1438
Db 1381 FONGKVKVFCQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWHQAIALRMEVLGCEAQDLY 1438

RESULT 2

US-10-047-257-1
; Sequence 1, Application US/10047257
; Publication No. US20020115152A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255.2
; CURRENT APPLICATION NUMBER: US/10/047,257
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-047-257-1

Query Match 100.0%; Score 7691; DB 14; Length 1438;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATRRYVLGAVELSWDYMQSDLGELPVDARPPRPVKSPFPNTSVVYKTKLFVEFTVHLFN 60
Db 1 ATRRYVLGAVELSWDYMQSDLGELPVDARPPRPVKSPFPNTSVVYKTKLFVEFTVHLFN 60
Qy 61 IAKPRPPMGLIGPTIOAEVYDVTVTILKNMASHPVSLHAGVSVYKASEGAEYDDQTSQ 120
Db 61 IAKPRPPMGLIGPTIOAEVYDVTVTILKNMASHPVSLHAGVSVYKASEGAEYDDQTSQ 120
Qy 121 REKEDKVPFGSGSHYVWQVLKENGPMASDPLCLTYSYLSHVLDLKDLSGLIGALLVCR 180
Db 121 REKEDKVPFGSGSHYVWQVLKENGPMASDPLCLTYSYLSHVLDLKDLSGLIGALLVCR 180
Qy 181 EGSIAKEKTTQTLHKETILLFAVDECKSWHSETKNSLMQDRDAASARAPKMTVNGYNR 240
Db 181 EGSIAKEKTTQTLHKETILLFAVDECKSWHSETKNSLMQDRDAASARAPKMTVNGYNR 240
Qy 241 SLPLGLIGCHRSVYWHVIGMGTTPPEVHSIFLFGHTFLVRNHRQASLEISPIFTLTAQTLL 300
Db 241 SLPLGLIGCHRSVYWHVIGMGTTPPEVHSIFLFGHTFLVRNHRQASLEISPIFTLTAQTLL 300
Qy 301 MDLGGFLFLFCHISSHQHDGMEAYVKVDCSPESPQLRMKNNEAEYDDDLTDSMDVVRP 360
Db 301 MDLGGFLFLFCHISSHQHDGMEAYVKVDCSPESPQLRMKNNEAEYDDDLTDSMDVVRP 360
Qy 361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLPDDRYSKYQYLANGPQIG 420
Db 361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLPDDRYSKYQYLANGPQIG 420
Qy 421 RYKVKVRFWAYTDETFKTRAIQHESGILGPLLYGEVGDITLLIIFKNQASRPYNIYPHGI 480
Db 421 RYKVKVRFWAYTDETFKTRAIQHESGILGPLLYGEVGDITLLIIFKNQASRPYNIYPHGI 480


```
Qy 481 TDVRLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGTPKSDPRCLTRYSSFNME 540
Db 481 TDVRLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGTPKSDPRCLTRYSSFNME 540
Qy 541 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDNRNRYLTENIQRLPNPAG 600
Db 541 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDNRNRYLTENIQRLPNPAG 600
Qy 601 VOLEDDPFOASNIHMSINGVYVDSLSQVCLHEVAYWYILSIGACTDPLSFVFFSGYTFKH 660
Db 601 VOLEDDPFOASNIHMSINGVYVDSLSQVCLHEVAYWYILSIGACTDPLSFVFFSGYTFKH 660
Qy 661 KMVYEDTLTLFPFSGEIVFMSMENPGLWILGCHNSDFNRGWTALLKVSSCDKNTGDIYE 720
Db 661 KMVYEDTLTLFPFSGEIVFMSMENPGLWILGCHNSDFNRGWTALLKVSSCDKNTGDIYE 720
Qy 721 DSYEDISAYLLSKNNAIBPRFSQNPVVKHQBREITRTTLOSQBEIDYDDTISVEMKK 780
Db 721 DSYEDISAYLLSKNNAIBPRFSQNPVVKHQBREITRTTLOSQBEIDYDDTISVEMKK 780
Qy 781 EDFDIYDEDNQSPRSFOKTRHYPIAAVERLWDYGMSSSPHVLNRNRAQSGSVQFKKV 840
Db 781 EDFDIYDEDNQSPRSFOKTRHYPIAAVERLWDYGMSSSPHVLNRNRAQSGSVQFKKV 840
Qy 841 FQETDGSFTQPLRGELNEHLGLGPYIRAEVEDNIMVTFNRNQAISFYSSLSIYEE 900
Db 841 FQETDGSFTQPLRGELNEHLGLGPYIRAEVEDNIMVTFNRNQAISFYSSLSIYEE 900
Qy 901 DORQABPRKQNFVFNETKTFWKVQHMAPTKDBFCCKAWAYSDVDLEKDVHSLGILP 960
Db 901 DORQABPRKQNFVFNETKTFWKVQHMAPTKDBFCCKAWAYSDVDLEKDVHSLGILP 960
Qy 961 LLVCHTNTLNPAHQROVTVQSFALFTTFIDETKSWYFTENMERNCRAPCNQIMEDPTPK 1020
Db 961 LLVCHTNTLNPAHQROVTVQSFALFTTFIDETKSWYFTENMERNCRAPCNQIMEDPTPK 1020
Qy 1021 NYRFAINGYIMDTLPGLVMAQDQRIWYLLSMGSNENIHSIFSGHYFTVRKKEEYKMA 1080
Db 1021 NYRFAINGYIMDTLPGLVMAQDQRIWYLLSMGSNENIHSIFSGHYFTVRKKEEYKMA 1080
Qy 1081 LYNYPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYNKCTPLGMAASHIRD 1140
Db 1081 LYNYPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYNKCTPLGMAASHIRD 1140
Qy 1141 FOITASQGYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQAROKPS 1200
Db 1141 FOITASQGYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQAROKPS 1200
Qy 1201 SLVYSQFTIMYSLDGKKWQYRGNSTGTLMVFFGNVDSGSIKHNIFNPPIIARYIRLHPT 1260
Db 1201 SLVYSQFTIMYSLDGKKWQYRGNSTGTLMVFFGNVDSGSIKHNIFNPPIIARYIRLHPT 1260
Qy 1261 HYSIRSTRMELMGLCDLNSCNPGLMESKASDAQITASSYFTNNFATWSPSKARLHQ 1320
Db 1261 HYSIRSTRMELMGLCDLNSCNPGLMESKASDAQITASSYFTNNFATWSPSKARLHQ 1320
Qy 1321 RSNARPOVNNPKWLQVDFOKTMKVTGVTQGVKSLTSMYKBEFLSSODGHOWTLF 1380
Db 1321 RSNARPOVNNPKWLQVDFOKTMKVTGVTQGVKSLTSMYKBEFLSSODGHOWTLF 1380
Qy 1381 FONGKVKVFGQNDQSFTPVNSLDPPLTRYLRHPQSVWVHQAIRMEVLGCEAQDLY 1438
Db 1381 FONGKVKVFGQNDQSFTPVNSLDPPLTRYLRHPQSVWVHQAIRMEVLGCEAQDLY 1438
```

RESULT 3

```
US-10-225-900-1
; Sequence 1, Application US/10225900
; Publication No. US2003007752A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
```

```
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255
; CURRENT APPLICATION NUMBER: US/10/225,900
; CURRENT FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: US/09/209,916
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-225-900-1
```

```
Query Match 100.0%; Score 7691; DB 15; Length 1438;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATRRYILGAVELSDWYMSDLGELPVDARPPRPVKSPFNTSVYKTLFVEFTVHLFN 60
Db 1 ATRRYILGAVELSDWYMSDLGELPVDARPPRPVKSPFNTSVYKTLFVEFTVHLFN 60
Qy 61 IAKPRPPWMLGLGPTIQAEVYDVITLKNWASHPSLHAGVSYWKASEGAEYDDQTSQ 120
Db 61 IAKPRPPWMLGLGPTIQAEVYDVITLKNWASHPSLHAGVSYWKASEGAEYDDQTSQ 120
Qy 121 REKEDDKVFPGSGHYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
Db 121 REKEDDKVFPGSGHYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
Qy 181 EGSIAKKTQTLHKFIILLFAVFDGKSWHSTKNSLMQDRDAASARAWPKHTNGYVNR 240
Db 181 EGSIAKKTQTLHKFIILLFAVFDGKSWHSTKNSLMQDRDAASARAWPKHTNGYVNR 240
Qy 241 SLPLGICHRKSVVHWVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTL 300
Db 241 SLPLGICHRKSVVHWVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTL 300
Qy 301 MDLQOFLFCHISSHQHDGMEAYVKVDSCEEPQLRMKNEEABDYDDDLTDSEMDVVR 360
Db 301 MDLQOFLFCHISSHQHDGMEAYVKVDSCEEPQLRMKNEEABDYDDDLTDSEMDVVR 360
Qy 361 DDNSPSFIQIRSVAKGHPKWTWVHYIAAEEDWDYAPLVLPADDRSYKSOYLNNGPQIR 420
Db 361 DDNSPSFIQIRSVAKGHPKWTWVHYIAAEEDWDYAPLVLPADDRSYKSOYLNNGPQIR 420
Qy 421 RYKVKVRFMAYTDETFKTRAIQHSGLGLPGLLYGEVDTLLIIFKNQASRPYNIYPHI 480
Db 421 RYKVKVRFMAYTDETFKTRAIQHSGLGLPGLLYGEVDTLLIIFKNQASRPYNIYPHI 480
Qy 481 TDVRLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGTPKSDPRCLTRYSSFNME 540
Db 481 TDVRLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGTPKSDPRCLTRYSSFNME 540
Qy 541 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDNRNRYLTENIQRLPNPAG 600
Db 541 RDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDNRNRYLTENIQRLPNPAG 600
Qy 601 VOLEDDPFOASNIHMSINGVYVDSLSQVCLHEVAYWYILSIGACTDPLSFVFFSGYTFKH 660
Db 601 VOLEDDPFOASNIHMSINGVYVDSLSQVCLHEVAYWYILSIGACTDPLSFVFFSGYTFKH 660
Qy 661 KMVYEDTLTLFPFSGEIVFMSMENPGLWILGCHNSDFNRGWTALLKVSSCDKNTGDIYE 720
Db 661 KMVYEDTLTLFPFSGEIVFMSMENPGLWILGCHNSDFNRGWTALLKVSSCDKNTGDIYE 720
Qy 721 DSYEDISAYLLSKNNAIBPRFSQNPVVKHQBREITRTTLOSQBEIDYDDTISVEMKK 780
Db 721 DSYEDISAYLLSKNNAIBPRFSQNPVVKHQBREITRTTLOSQBEIDYDDTISVEMKK 780
```

```
Db 721 DSYEDISAYLSKNNAIIPRSFSQNPVVKRQREITRTTLQSDQEEIDYDDTISVEMKK 780
Qy 781 EDYDIDEDENQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGSVQFKKVV 840
Db 781 EDYDIDEDENQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGSVQFKKVV 840
Qy 841 FQEFDTGDSFTQPLRGELNEHLGLGPYIRAEVDNIMVTFRNQASRPYSYSSLSIYEE 900
Db 841 FQEFDTGDSFTQPLRGELNEHLGLGPYIRAEVDNIMVTFRNQASRPYSYSSLSIYEE 900
Qy 901 DORQAEPRKQFVKENETKTYFWKQHEMAFKDEFCKAKAYSDVDLEKDVHSGLGP 960
Db 901 DORQAEPRKQFVKENETKTYFWKQHEMAFKDEFCKAKAYSDVDLEKDVHSGLGP 960
Qy 961 LLVCHTNTLNPAHQGVTVQSFALPFTTFDETSWYFTENMERNCRACNTQMEDPTKE 1020
Db 961 LLVCHTNTLNPAHQGVTVQSFALPFTTFDETSWYFTENMERNCRACNTQMEDPTKE 1020
Qy 1021 NYRFHAINGYIMDTLPLGLVMAQDQRIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKQA 1080
Db 1021 NYRFHAINGYIMDTLPLGLVMAQDQRIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKQA 1080
Qy 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCQTPLGMASGHIRD 1140
Db 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCQTPLGMASGHIRD 1140
Qy 1141 FOITASGOYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGKIKTOGAROKFS 1200
Db 1141 FOITASGOYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGKIKTOGAROKFS 1200
Qy 1201 SLIYISQFIIMYSLDGKQWYRGNSTGTLMWFFGNVSSGIKHNIIPNPPIIARYIRLHPT 1260
Db 1201 SLIYISQFIIMYSLDGKQWYRGNSTGTLMWFFGNVSSGIKHNIIPNPPIIARYIRLHPT 1260
Qy 1261 HYSIRSTRLMELMGCDLNSCMPLGESKAIISDAQITASSYFTNMFATWSPSKARLHLQ 1320
Db 1261 HYSIRSTRLMELMGCDLNSCMPLGESKAIISDAQITASSYFTNMFATWSPSKARLHLQ 1320
Qy 1321 RSNARWQVNNPKEWLQVDFQTKMKTGVTTOGVKSLLTSMVVKGFLLSSSDQGHQWTLF 1380
Db 1321 RSNARWQVNNPKEWLQVDFQTKMKTGVTTOGVKSLLTSMVVKGFLLSSSDQGHQWTLF 1380
Qy 1381 FQNGKVKVQFGNQDSFTPVVNSLDPLLLTRYLRHQPQSWHQAIALRMEVLGCEAQDLY 1438
Db 1381 FQNGKVKVQFGNQDSFTPVVNSLDPLLLTRYLRHQPQSWHQAIALRMEVLGCEAQDLY 1438
```

RESULT 4

US-10-095-718-2

; Sequence 2, Application US/10095718

; Publication No. US20020131956A1

; GENERAL INFORMATION:

; APPLICANT: Walsh, Christopher

; APPLICANT: Chao, Hengjun

; APPLICANT: Burstein, Haim

; APPLICANT: Lynch, Carmel

; APPLICANT: Stepan, Tony

; APPLICANT: Munson, Keith

; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and

; TITLE OF INVENTION: Methods of Using the Same

; FILE REFERENCE: 35052/204375

; CURRENT FILING DATE: US/10/095,718

; PRIOR FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: 09/689,430

; PRIOR FILING DATE: 2001-08-22

; PRIOR APPLICATION NUMBER: 60/158,780

; PRIOR FILING DATE: 1999-10-12

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 1471

; TYPE: PRT

; ORGANISM: Homo sapiens B-domain deleted factor VIII

```
; FEATURE:
; OTHER INFORMATION: Homo sapiens BDD FVIII
US-10-095-718-2
```

```
Query Match 99.8%; Score 7674; DB 14; Length 1471;
Best Local Similarity 99.08; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 14; Gaps 1;
```

```
Qy 1 ATRYYILGAVELSDWYMQSDLGELPVDARPPRPVPSKSPFPNTSVVYKTLFVEPTVHLFN 60
Db 20 ATRYYILGAVELSDWYMQSDLGELPVDARPPRPVPSKSPFPNTSVVYKTLFVEPTVHLFN 79
Qy 61 IAKRPPWMLGILGPTIOAEVYDVTWITLKNASHPVSILHAGVSYWKAEGAEVDDQTSQ 120
Db 80 IAKRPPWMLGILGPTIOAEVYDVTWITLKNASHPVSILHAGVSYWKAEGAEVDDQTSQ 139
Qy 121 REKEDDKVFPGGSHYYQVQLKENGPMASDPLCLITYSLSHVDLVKDLNLSLIGALLVCR 180
Db 140 REKEDDKVFPGGSHYYQVQLKENGPMASDPLCLITYSLSHVDLVKDLNLSLIGALLVCR 199
Qy 181 EGSIAKEKTQTLHKFILLFAVDFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
Db 200 EGSIAKEKTQTLHKFILLFAVDFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 259
Qy 241 SLPLGILGCHRSVYWHVIGMGTTFEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 300
Db 260 SLPLGILGCHRSVYWHVIGMGTTFEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 319
Qy 301 MDLQGFLLFCHISSHQHDGMEAYVKVDSCPEEPOLRMKNNEAEYDDDLTDSMDVYVRF 360
Db 320 MDLQGFLLFCHISSHQHDGMEAYVKVDSCPEEPOLRMKNNEAEYDDDLTDSMDVYVRF 379
Qy 361 DDDNSPSFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLVLAPDPRSYKQYLNNQPORG 420
Db 380 DDDNSPSFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLVLAPDPRSYKQYLNNQPORG 439
Qy 421 RYKVKRPMAYTDEFTKREAIQHESGILGLPGLGEGVDTLLIIFKQOASPYNIYPHGI 480
Db 440 RYKVKRPMAYTDEFTKREAIQHESGILGLPGLGEGVDTLLIIFKQOASPYNIYPHGI 499
Qy 481 TDVRLYRRRLPKGVKHLKDPFLLPGEIFPKYKWTVYVEDGPKTSDPRCLTYYSFVNME 540
Db 500 TDVRLYRRRLPKGVKHLKDPFLLPGEIFPKYKWTVYVEDGPKTSDPRCLTYYSFVNME 559
Qy 541 RDLASGLIGPLLCYKESVDQRGNQIMSDKRNVLFSVFDENRSHYLTENIQRPINPAG 600
Db 560 RDLASGLIGPLLCYKESVDQRGNQIMSDKRNVLFSVFDENRSHYLTENIQRPINPAG 619
Qy 601 VQLEDPEFQASNMHSGINGYVFDLSQLSVCLHEVAYWYILSGAGTDFLSVFFSGYTKH 660
Db 620 VQLEDPEFQASNMHSGINGYVFDLSQLSVCLHEVAYWYILSGAGTDFLSVFFSGYTKH 679
Qy 661 KMVYEDTTLTPPFSGETVFMSENPLWLILGCHNSDFNRGMTALLKYSSCDKNTGDIYE 720
Db 680 KMVYEDTTLTPPFSGETVFMSENPLWLILGCHNSDFNRGMTALLKYSSCDKNTGDIYE 739
Qy 721 DSYEDISAYLSKNNAIIPRSFSQNPVVKRQREITRTTLQSDQEEIDYDDTISVEMKK 766
Db 740 DSYEDISAYLSKNNAIIPRSFSQNPVVKRQREITRTTLQSDQEEIDYDDTISVEMKK 799
Qy 767 EIDYDDTISVEMKKEDFDIYDENQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR 826
Db 800 EIDYDDTISVEMKKEDFDIYDENQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR 859
Qy 827 RAQSGSVQFKKVFQEFDTGDSFTQPLRGELNEHLGLGPYIRAEVDNIMVTFRNQAS 886
Db 860 RAQSGSVQFKKVFQEFDTGDSFTQPLRGELNEHLGLGPYIRAEVDNIMVTFRNQAS 919
Qy 887 RPYSFYSSLSIYEEEDQROGAEPKRNKVPKFNKTKTYFWKQHEMAFKDEFCKAKAYFSD 946
Db 920 RPYSFYSSLSIYEEEDQROGAEPKRNKVPKFNKTKTYFWKQHEMAFKDEFCKAKAYFSD .979
Qy 947 VDLEKDVHSGLIGLPLLVCHTNTLNPAHQGVTVQSFALPFTTFDETSWYFTENMERNCR 1006
```

Db 980 VLEKDVHSLGLPLLVCHTNTLNPAGRQVTVQSFALFTTIFDETKSWYFTENMERNCR 1039
QY 1007 APCNIQMEPTFKENYRFHAINGYIMDTLPGLVMAQDQIRWYLLSGNSNENIHSFHFG 1066
Db 1040 APCNIQMEPTFKENYRFHAINGYIMDTLPGLVMAQDQIRWYLLSGNSNENIHSFHFG 1099
QY 1067 HVTVRKKEBYKALYNLYPGVPEVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNK 1126
Db 1100 HVTVRKKEBYKALYNLYPGVPEVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNK 1159
QY 1127 CQPLGASHIIRDFQITASGOYQOWAPKARLHYSGSINAWSTKPPFWIKVDLLAPMI 1186
Db 1160 CQPLGASHIIRDFQITASGOYQOWAPKARLHYSGSINAWSTKPPFWIKVDLLAPMI 1219
QY 1187 IHGKITQAROKFSSLYISQFIIMYSLDGKKWOTYRGNSGTLMVFFGNVDSGGIKHNIF 1246
Db 1220 IHGKITQAROKFSSLYISQFIIMYSLDGKKWOTYRGNSGTLMVFFGNVDSGGIKHNIF 1279
QY 1247 NPPIIARYIRLPHTHYSIRSLRMELMGCDLNSCMLPGMESKAISDAQITASSYFTNMF 1306
Db 1280 NPPIIARYIRLPHTHYSIRSLRMELMGCDLNSCMLPGMESKAISDAQITASSYFTNMF 1339
QY 1307 ATWSPSKARLHLQGRSNAMRPQVNNPKWLOVDFOKTMKVTGVTTOGVKSLTSMYKBF 1366
Db 1340 ATWSPSKARLHLQGRSNAMRPQVNNPKWLOVDFOKTMKVTGVTTOGVKSLTSMYKBF 1399
QY 1367 L1SSSDGHWTLFFQNGKVKVQGNQDSFTPVVNSLDPLLRILYRIHPQSWHQAIALR 1426
Db 1400 L1SSSDGHWTLFFQNGKVKVQGNQDSFTPVVNSLDPLLRILYRIHPQSWHQAIALR 1459
QY 1427 MEVLGCEAODLY 1438
Db 1460 MEVLGCEAODLY 1471

RESULT 5

US-09-957-641-2
; Sequence 2, Application US/09957641
; Publication No. US20020182670A1
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-00
; CURRENT APPLICATION NUMBER: US/09/957,641
; CURRENT FILING DATE: 2001-09-16
; PRIOR APPLICATION NUMBER: US 60/234047
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/236460
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-957-641-2

Query Match 94.1%; Score 7234; DB 10; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
QY 1 ATRRYIAGVELSWDMQSLGELPVDARPPRPVKSPFPNTSVVYKKTLPVEFTVHLEN 60
Db 1 ATRRYIAGVELSWDMQSLGELPVDARPPRPVKSPFPNTSVVYKKTLPVEFTVHLEN 60
QY 61 IAKPRPWWGLLGTPTQAEYVDYVITLKNWASHPVSLHAGVSYWKASGAEYDDQTSQ 120
Db 61 IAKPRPWWGLLGTPTQAEYVDYVITLKNWASHPVSLHAGVSYWKASGAEYDDQTSQ 120
QY 121 REKEDKVPFGGSHYVQVQLKNGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
Db 121 REKEDKVPFGGSHYVQVQLKNGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180

QY 181 EGS LAKEKKTQTLHKFTILLFAVFDGSKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
Db 181 EGS LAKEKKTQTLHKFTILLFAVFDGSKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
QY 241 SLPGILGCHRKSVYHVICMGTTPEVHSIFLAGHTEFLVNHROASLEISPIITFLTAQTLL 300
Db 241 SLPGILGCHRKSVYHVICMGTTPEVHSIFLAGHTEFLVNHROASLEISPIITFLTAQTLL 300
QY 301 MDLQGFLLFCHTSSHOHQMEAYVKVDSCEPPEPQLRMKNNEAEADYDDDLTSEMDVVF 360
Db 301 MDLQGFLLFCHTSSHOHQMEAYVKVDSCEPPEPQLRMKNNEAEADYDDDLTSEMDVVF 360
QY 361 DDNSPSFQIRSVAKCKHPTKTVHYIAAEEEDWDYAPLVLAPDDRYSKYQYLANGPQIRG 420
Db 361 DDNSPSFQIRSVAKCKHPTKTVHYIAAEEEDWDYAPLVLAPDDRYSKYQYLANGPQIRG 420
QY 421 RYKVKYRFMAYTDETFKTRTALQHESGILGPLLYGVEGDTLLIIFKNQASRPYNIYPHGI 480
Db 421 RYKVKYRFMAYTDETFKTRTALQHESGILGPLLYGVEGDTLLIIFKNQASRPYNIYPHGI 480
QY 481 TDVRPLYSRRLPGVVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSFVNME 540
Db 481 TDVRPLYSRRLPGVVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSFVNME 540
QY 541 ROLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVDFENRSWYLTENIQRFLENPAG 600
Db 541 ROLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVDFENRSWYLTENIQRFLENPAG 600
QY 601 VQLEDPFOASNIMHSINGYVFDLSQLSVCLHEVAWYIILSGAQTDFLSVPSFGYTFKH 660
Db 601 VQLEDPFOASNIMHSINGYVFDLSQLSVCLHEVAWYIILSGAQTDFLSVPSFGYTFKH 660
QY 661 KMVYEDTLTLFPFSGETVFMENPGLMILGCHNSDFRNKGMTALLKVKSCDKNTGDYIE 720
Db 661 KMVYEDTLTLFPFSGETVFMENPGLMILGCHNSDFRNKGMTALLKVKSCDKNTGDYIE 720
QY 721 DSYEDISAYLLSKNNAIEPRSF 742
Db 721 DSYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTIPENDIEKTDPMFAHRTMP 780
QY 743 742
Db 781 KIQNVSSDLMLLRQSPHGLSLDLQEAQYETFDSPGAISSNNLSSEMTFRPQ 840
QY 743 742
Db 841 LHHSGDMVTPPSGLQLRNEKLGTTAATLKKLDPKVSTSNLSTIPSDNLAAGTN 900
QY 743 742
Db 901 TSSLGPPMPVHYDQSLDTTLFGKKSSPLTESGGPLSLEENNDKLLSGLMNSQESSW 960
QY 743 742
Db 961 GKNVSSSTESGLFKGKRAHPALLTKDNALFKVSIISLLKTNKTSNNSATNRKTHIDGPSL 1020
QY 743 742
Db 1021 LIENSPPVQNTLESDETEPKKVTPLIHDRMLMDKNATALLNHNMSKNTSSKNMVMVQK 1080
QY 743 742
Db 1081 KEGPIPPDAQNDPMSFFKMLFLPESARWIORTHGKNSLNSGQGPSKQLVSLGPEKSV 1140
QY 743 742
Db 1141 QNPLSEKNVVKVKGSETKDVGKEMVFPSSRNLFUTNLDNLHNHNTHNQEKIQEIEK 1200
QY 743 742
Db 1201 KETLIOENVVLPIQIHTVTGTGKFMKNLFLLSSTRQNVESYEGAYAPVLQDFRSLNDS 1260

Qy 743 ----- 742
Db 1261 TKHTAHFSKKGBEENLEGLNQTQKQIVEKYACTTRISPNTSQNFVQTSKRALKQFRL 1320
Qy 743 ----- 742
Db 1321 PLEETELEKRIIVDDTSTQSKNMKHLTPSTLTQIDYNEKEKGAITQSPLDCLTRSHSI 1380
Qy 743 ----- 742
Db 1391 PQANRSPPIAKVSPFPIRPIYLRVLFDQNSSHLPAASYRKXKDSGVQBSHFLQGA 1440
Qy 743 ----- 742
Db 1441 NNLSLAILLTLEMTGDQREVGSIGTSATNSVTYKVENTVLKPKPOLPKTSGKVELLPKVI 1500
Qy 743 ----- 742
Db 1501 YOKOLFPTTETNGSPGHLDLVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESSAKTPSK 1560
Qy 743 ----- 742
Db 1561 LLDPLANDHNYGTQPKBEWKSQKSPKTAFFKKDDTILSNACSNHAIANEGQNP 1620
Qy 743 ----- 742
Db 1621 EIEVTWAKQRTERLCSQNPVVKHREITRTTLOSQDEEIDYDDTISVENKKEDFDIY 1680
Qy 743 ----- 742
Db 1681 DEDENQSPRSQKTRHYFIAAVERLWDYGMSSSHVLRNRAQSGSVQPKVQFOETD 1740
Qy 743 ----- 742
Db 1741 GSFTQPLVARGELNEHLGGLGPVIRAEVDNIMVTRFNQASRPSYSSLSIYEEQORQA 1800
Qy 743 ----- 742
Db 1801 EPRKNFVKNETKTYFWKQVHMAPTKDFCKAWAYFSDVLEKDVHSGILGLLVCHT 1860
Qy 743 ----- 742
Db 1861 NTLNPAHQVQVQSFALFPTTIDETKSWYFTFENMERNCRAPCNQMEDPTTKENYRPHA 1920
Qy 743 ----- 742
Db 1921 INGYIMDTLPGVMAQDQIRWYLLSMGSENIHSHFGSHVFTVRKKEEYKALYNLYP 1980
Qy 743 ----- 742
Db 1981 GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGMAHGHRDFOITAS 2040
Qy 743 ----- 742
Db 2041 GOYGOWAPKLARLHYSGSNAMSTKEPFWIKVDLLAPMIHGIKTQGARQKFSLSYISQ 2100
Qy 743 ----- 742
Db 2101 FIIMYSLDGKWKQTYRGNSTGLTWFFGNVDSSGICKHNFNPPIIARYIRLHPHYIRS 2160
Qy 743 ----- 742
Db 2161 TLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMPATWSPSKARLHLQGRNMR 2220
Qy 743 ----- 742
Db 2221 PQVNNPKWLOVDFOKTMKVTGTTQGVKSLLSMVKPELISSSQDGHQWTLFPONGKV 2280
Qy 743 ----- 742
Db 2281 KVFQGNQDSFTPVVNSLDPPLLTRYLRHQPQSWHQAIALRMEVLGCEAODLY 1438
2281 KVFQGNQDSFTPVVNSLDPPLLTRYLRHQPQSWHQAIALRMEVLGCEAODLY 2332

RESULT 6

US-10-131-510A-2
; Sequence 2, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; NUMBER OF SEQ ID NOS: 40
; SEQ ID NO 2
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-510A-2

Query Match 94.1%; Score 7234; DB 12; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
Qy 1 ATRRYLCAVELSWDMQSDLGELPVDARFPVPKSPPTNTSVVYKKTLPVEFTVHLFN 60
Db 1 ATRRYLCAVELSWDMQSDLGELPVDARFPVPKSPPTNTSVVYKKTLPVEFTVHLFN 60
Qy 61 IAKRPPPMWGLLGPTTQAEVVDVTVTILKNMASHPVSLHVGVSYWKASEGAEYDDQTSQ 120
Db 61 IAKRPPPMWGLLGPTTQAEVVDVTVTILKNMASHPVSLHVGVSYWKASEGAEYDDQTSQ 120
Qy 121 REKEDDKVPFGSHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
Db 121 REKEDDKVPFGSHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
Qy 181 EGSLAKEKTOTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAMPKQHTVNGYVNR 240
Db 181 EGSLAKEKTOTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAMPKQHTVNGYVNR 240
Qy 241 SLPLGIGCHRSVYWHVIGMGTTPRVHSIFLEGHTFLVNRHQASLEISPTIFLTAQTLL 300
Db 241 SLPLGIGCHRSVYWHVIGMGTTPRVHSIFLEGHTFLVNRHQASLEISPTIFLTAQTLL 300
Qy 301 MDLQGFLLFCHTSSHQHGMAYKVDSCPEPQLRMKNNEAEYDDDLTDTSEMDVVRP 360
Db 301 MDLQGFLLFCHTSSHQHGMAYKVDSCPEPQLRMKNNEAEYDDDLTDTSEMDVVRP 360
Qy 361 DDNSPSPFIQIRSVAKKHPKTTWHVIAAEEEDWDYAPLVADDDRYSKSOYLNNGPQIR 420
Db 361 DDNSPSPFIQIRSVAKKHPKTTWHVIAAEEEDWDYAPLVADDDRYSKSOYLNNGPQIR 420
Qy 421 RYKVKVRFMAYTDTFTKREAIQHSGLIGPLLYGEVGDTLIIIFKNQASRPNTYPHGI 480
Db 421 RYKVKVRFMAYTDTFTKREAIQHSGLIGPLLYGEVGDTLIIIFKNQASRPNTYPHGI 480
Qy 481 TDVRLYLSRRLPKGVKHLKDPFIPGRIPKYKWTVTVEDGPTKSDPRCLTRYSSFFVNM 540
Db 481 TDVRLYLSRRLPKGVKHLKDPFIPGRIPKYKWTVTVEDGPTKSDPRCLTRYSSFFVNM 540
Qy 541 RDLASGLIGPLLIICYKESVDQRGNQIMSDKRVNVLFSVFDENRSWLTENIQRFLEPNAG 600

```
Db 541 RDLASGLIGLLICYSVDRGQINSDKRNVLFSVFDENRSWYLTENIRFLPNPAG 600
Qy 601 VQLEDPFQASINMHSINGVDFSLQLSVCLHEVAYWYLLSIGAQTDPLSVFFSGYTFKH 660
Db 601 VQLEDPFQASINMHSINGVDFSLQLSVCLHEVAYWYLLSIGAQTDPLSVFFSGYTFKH 660
Qy 661 KMVEDTLTLFPFSGETVFMSEMPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDYIE 720
Db 661 KMVEDTLTLFPFSGETVFMSEMPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDYIE 720
Qy 721 DSYEDISAYLLSKNNAIEPRSP-----742
Db 721 DSYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTIPENDIBKTDPFWFAHRTMP 780
Qy 743 -----742
Db 781 KIQNVSSDLMLLRQSPTHGLSLSDLOQAKYETTFSDDSPGALDSNNSISEMTHFRPQ 840
Qy 743 -----742
Db 841 LHHSGDMVFTPESGLOLRLEKLGTTAATLKKLDKFKVST'SNNLISTIPSDNLAAGTDN 900
Qy 743 -----742
Db 901 TSSLGPPSPVHYDSQDSTLTFGKKSPLETESGGPLSLSEENNDKLLSGLMNSQSSW 960
Qy 743 -----742
Db 961 GKNVSTESGELFKGKAHGPALLTKONALFKVSIILLKTKNTSNNSATNRKTHIDGPSL 1020
Qy 743 -----742
Db 1021 LIENSPSWQWNILESDETFKKVTPLIHDRMLMDKNATALRLNHSNNTTSSKNMEMYQOK 1080
Qy 743 -----742
Db 1081 KEGFIPDPAQNPDMSFFKMLFLPESARWIQTHGKNSLNSGQSPKQLVSLGPEKSVEG 1140
Qy 743 -----742
Db 1141 QNFLSEKNVVGGEFTKDVGLKEMVFPSSRNLI'FLNLDNLHENNTHNQEKIQEIEK 1200
Qy 743 -----742
Db 1201 KETLIQENVVLPQIHVTGTGKNFMKNLFLILSTRQNVESYEGAYAPVLQDFRSLNDSTNR 1260
Qy 743 -----742
Db 1261 TKKHTAHFSKGBEENLEGLNOTKQIYEKYACTTRISPNTSQOQFVTVQSRKALKQFRL 1320
Qy 743 -----742
Db 1321 PLBETELEKRIIVDDTSTQSKMKHUTPSTLTQIDYNEKEGAITQSPSLDCLTRSHSI 1380
Qy 743 -----742
Db 1381 PQANRSPUIAKVSSPFSIRPIYLRVLFDQNSSHLPAASVRKDSGVQESSHFLQAKK 1440
Qy 743 -----742
Db 1441 NNLSLAILTLEMTGDQREVSLGTSATNSVTYKKVENTVLPKPLDKPTSGKVELLPKVIH 1500
Qy 743 -----742
Db 1501 YQKDLPTTETNSGPHLDIVESLLOGTEGAIKWNEANRPGKVPFLRVATESAKTPSK 1560
Qy 743 -----742
Db 1561 LLDPLANDNHYGTQIPKEEWSQESPEKTAFFKKOTILSNACESNHAATAINEGONKP 1620
Qy 743 -----SQNPVPLKRRHQREITRRTLQSDQBEIDYDDTISVEMKEDFDIY 786
|||||
```

```
Db 1621 EIEVTWAKQRTERLCSQNPVPLKRRHQREITRRTLQSDQBEIDYDDTISVEMKEDFDIY 1680
Qy 787 DEBENQSPRSFQKTRHYFIAAVERLWDYQWSSSPHVLNRNRAQSGSVQPKKVVFOBFTD 846
Db 1681 DEBENQSPRSFQKTRHYFIAAVERLWDYQWSSSPHVLNRNRAQSGSVQPKKVVFOBFTD 1740
Qy 847 GSTQPIYRGELNEHLLGLGPYIRAEVEDNIMVTFRNOASRPYSFYSSLSISYEDDQROGA 906
Db 1741 GSTQPIYRGELNEHLLGLGPYIRAEVEDNIMVTFRNOASRPYSFYSSLSISYEDDQROGA 1800
Qy 907 EPRKNFVKNETKTYFWKVOHMAPTKDBFCDAWAYFSDVDLEKOVHSLIGPLLCHT 966
Db 1801 EPRKNFVKNETKTYFWKVOHMAPTKDBFCDAWAYFSDVDLEKOVHSLIGPLLCHT 1860
Qy 967 NTLNPAHGRQVTVQEFALFETIFDETKSWYFTENMERNCRAPCNIQMEDTFFKENYRPHA 1026
Db 1861 NTLNPAHGRQVTVQEFALFETIFDETKSWYFTENMERNCRAPCNIQMEDTFFKENYRPHA 1920
Qy 1027 INGYIMDTLPLGVMAQDQIRWYLLSMGSENTHSIHFSGHVPTVRKKEBKALYNLYP 1086
Db 1921 INGYIMDTLPLGVMAQDQIRWYLLSMGSENTHSIHFSGHVPTVRKKEBKALYNLYP 1980
Qy 1087 GVFEVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGASGHIRDFOITAS 1146
Db 1981 GVFEVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGASGHIRDFOITAS 2040
Qy 1147 GOYQWAPKLARLHYSGSINAWSTKBPFSWKIVDLPAMIHIGIKTQCARQKFSLLYISQ 1206
Db 2041 GOYQWAPKLARLHYSGSINAWSTKBPFSWKIVDLPAMIHIGIKTQCARQKFSLLYISQ 2100
Qy 1207 FIIMYSLDGKKWQTYRGNSTGTLWVFFGNVDSGIXHNIENPFIIRYIRLHPHYSIRS 1266
Db 2101 FIIMYSLDGKKWQTYRGNSTGTLWVFFGNVDSGIXHNIENPFIIRYIRLHPHYSIRS 2160
Qy 1267 TLRWELMGCDLNSCSPHGLMESKAISDAQITASSYFTNMFATWSPSKARLHLOGRSNAWR 1326
Db 2161 TLRWELMGCDLNSCSPHGLMESKAISDAQITASSYFTNMFATWSPSKARLHLOGRSNAWR 2220
Qy 1327 POVNNPKWLVQDFQKTMKVTVTQGVKSLTSMYVKEFLISSODGHQWTLFFQNGKV 1386
Db 2221 POVNNPKWLVQDFQKTMKVTVTQGVKSLTSMYVKEFLISSODGHQWTLFFQNGKV 2280
Qy 1387 KVFQGNQDSPTPVVNSLDPPLRLYRIHPQSWHQAIALRMEVLGCEAODLY 1438
Db 2281 KVFQGNQDSPTPVVNSLDPPLRLYRIHPQSWHQAIALRMEVLGCEAODLY 2332
```

RESULT 7

US-10-187-319-2

; Sequence 2, Application US/10187319

; Publication No. US20030068785A1

; GENERAL INFORMATION:

; APPLICANT: Lollar, John S.

; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII

; NUMBER OF SEQUENCES: 40

; CORRESPONDENCE ADDRESS:

; ADDRESSES: Greenlee, Winner and Sullivan, P.C.

; STREET: 5370 Manhattan Circle Suite 201

; CITY: Boulder

; STATE: Colorado

; COUNTRY: USA

; ZIP: 80303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/187,319

; FILING DATE: 27-Aug-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 09/523,656

FILED DATE: 2000-03-10
APPLICATION NUMBER: US 09/037,601
FILING DATE: 1998-03-10
APPLICATION NUMBER: WO PCT/US97/11155
FILING DATE: 1997-06-26
APPLICATION NUMBER: US 08/670,707
FILING DATE: 1996-06-26
ATTORNEY/AGENT INFORMATION:
NAME: Greenlee, Lorraine L.
REGISTRATION NUMBER: 27,894
REFERENCE/DOCKET NUMBER: 75-95K
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: <Unknown>
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Liver
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-187-319-2

Query Match 94.1%; Score 7234; DB 15; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
QY 1 ATRRYLGAVELSWDMSQDLGELPVDAREPPVPKSPFNTSVVYKTLFVEFTVHLFN 60
DB 1 ATRRYLGAVELSWDMSQDLGELPVDAREPPVPKSPFNTSVVYKTLFVEFTVHLFN 60
QY 61 IAKPRPPMGLLGTQAEVYDVITLKNMASHPVSLHAGVSYWKASEGAEYDDQTSQ 120
DB 61 IAKPRPPMGLLGTQAEVYDVITLKNMASHPVSLHAGVSYWKASEGAEYDDQTSQ 120
QY 121 REKEDKVPFGSGHTYVWQVLEKNGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
DB 121 REKEDKVPFGSGHTYVWQVLEKNGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
QY 181 EGSIAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
DB 181 EGSIAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
QY 241 SLPLGLICHRKSVVWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTL 300
DB 241 SLPLGLICHRKSVVWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTL 300
QY 301 MDLQQLFLFCHISHQHDGMEAYVKVDSCPEEPQLRMKNNEAEYDDDLTDSMOWVRF 360
DB 301 MDLQQLFLFCHISHQHDGMEAYVKVDSCPEEPQLRMKNNEAEYDDDLTDSMOWVRF 360
QY 361 DDNSPSFIQIRSVAKGPKTWHYIAAEEDWDYAPLVLPDDRYSYKSOYLNNGPORIG 420
DB 361 DDNSPSFIQIRSVAKGPKTWHYIAAEEDWDYAPLVLPDDRYSYKSOYLNNGPORIG 420
QY 421 RYKVKRFMAVDTDETKREAIQHESGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHGI 480
DB 421 RYKVKRFMAVDTDETKREAIQHESGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHGI 480
QY 481 TDVRLPSRRLPKGVKHLKOPILLPGEIFKYKWTVTVEDGPKSDPRCLTRYSSFYVME 540
DB 481 TDVRLPSRRLPKGVKHLKOPILLPGEIFKYKWTVTVEDGPKSDPRCLTRYSSFYVME 540
QY 541 RDLASGLIGPLLIICYKESVDQRGNQIMSDKKNVILFVSFDENRNSWYLTENIQRELNPAG 600
DB 541 RDLASGLIGPLLIICYKESVDQRGNQIMSDKKNVILFVSFDENRNSWYLTENIQRELNPAG 600

QY 601 VQLEDPEFOASNIMHSINGYVDFSLQLSVCLHEVAYWILSIGAQTDLSVFFSGYTPKH 660
DB 601 VQLEDPEFOASNIMHSINGYVDFSLQLSVCLHEVAYWILSIGAQTDLSVFFSGYTPKH 660
QY 661 KMVYEDTLTLFPFSGETVFMSENPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDYYE 720
DB 661 KMVYEDTLTLFPFSGETVFMSENPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDYYE 720
QY 721 DSYEDISAYLLSKNNAIEPRSF----- 742
DB 721 DSYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTIPENDIEXTDWFAPHTPMP 780
QY 743 ----- 742
DB 781 KIQNVSSDLLMLLRQSPTPHGLSLSLDLQEAKEYTFSDDPSPGAIDSNNSISEMTHFRPQ 840
QY 743 ----- 742
DB 841 LHHSGDMVTPESGLQLRLNEKLGHATAATELKKLDFKVSSSTSNLLISTIPSDNLAAGTDN 900
QY 743 ----- 742
DB 901 TSSLGPSPMPVHYDSQLDTTLFGKSSPLTESGGPLSLSEENNDKLLSGLMNSQESSW 960
QY 743 ----- 742
DB 961 GKNVSTESGRLFKGRAGPALTKDNALFKVSI SLLKTNKTSNNSATNRKTHIDGPSL 1020
QY 743 ----- 742
DB 1021 LIENSFVWQNIESTDETFKVTPLIHDRMLMDKNATALLNHNMSKNTTSSKNMENVQOK 1080
QY 743 ----- 742
DB 1081 KEGPIPPDAQNDMSFFKMLFLPESARWITQTHGKNSLNSGQSPSPKQLVSLGPEKSVBG 1140
QY 743 ----- 742
DB 1141 QNFLSEKNKVVVGKGF TKDVLGKEMVFPSSRNLFITNLDNLHENNTHNQEKKIOBEIEK 1200
QY 743 ----- 742
DB 1201 KETLIQENVVLQIHTVTGKFMKMLFLLLSTRQNVGSEYAGAYAPVLQDFRSLNDSTR 1260
QY 743 ----- 742
DB 1261 TKKHTAHSKGBEENLEGNGQTQI1VEKYACTTRI SPNTSQNFVTOQRKRALQFRL 1320
QY 743 ----- 742
DB 1321 PLEETELEKRIIIVDDTSTQSKNMKHLTPSTLTQIDYNEKKGAIQTOSPLSDCLTRSHSI 1380
QY 743 ----- 742
DB 1381 PQANRSPPIAKVSSPSPRIPIYLTVLFDQNSSHLPAASRYKKDSGVQESSHFLOGAKK 1440
QY 743 ----- 742
DB 1441 NNLSLAILTLEMTGDQREVGLTSATNSVYKVKVENTVLPKPDLPKTSKGVKELLPKVHI 1500
QY 743 ----- 742
DB 1501 YQKDLFPFTETSNPGHLDLVEGSLAQGTGAIKMNEANRPGKVPFLRVATESSAKTPSK 1560
QY 743 ----- 742
DB 1561 LLDPLAWDNHYGTQIPKEBWKSEKSPKTAFFKKDTILSLNACENHAIANAENQONKP 1620
QY 743 ----- 786
DB 1621 EIEVTWAKQORTLCSQNPVLRKHQREITRTTLQSDQDEIDYDDTISVEMKEDFDIY 1680

QY 787 DEBENQSPRFOKTRHYFTAAVERLWDYGMSSSPHYLRNRAQSGSVPOKQVVFQEF 846
Db 1681 DEBENQSPRFOKTRHYFTAAVERLWDYGMSSSPHYLRNRAQSGSVPOKQVVFQEF 1740
QY 847 GSTFQPLRGELNEHLGLGPYIRAEVEDNIMWTFRNQASRPYSFYSSLLSYSEDORQGA 906
Db 1741 GSTFQPLRGELNEHLGLGPYIRAEVEDNIMWTFRNQASRPYSFYSSLLSYSEDORQGA 1800
QY 907 EPRKNFVKNETKTYFMKVQHMAPTKDEDFCKAWAYFSDVDLEKDVHSLGLIPLLVCHT 966
Db 1801 EPRKNFVKNETKTYFMKVQHMAPTKDEDFCKAWAYFSDVDLEKDVHSLGLIPLLVCHT 1860
QY 967 NTLNPAHGRQVTVQEFALFTTIDEKTSWYFTENMERNCRAPCNIMQEDPTFKENTRPHA 1026
Db 1861 NTLNPAHGRQVTVQEFALFTTIDEKTSWYFTENMERNCRAPCNIMQEDPTFKENTRPHA 1920
QY 1027 INGVIYMDTLPLGLVMAQDORIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKVALNLYP 1086
Db 1921 INGVIYMDTLPLGLVMAQDORIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKVALNLYP 1980
QY 1087 GVFTVEMPLPSKAGIMRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDFQITAS 1146
Db 1981 GVFTVEMPLPSKAGIMRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDFQITAS 2040
QY 1147 GOYGOWAPKARLHYSGSINAWSTKEPFSWIKVDLAPMIHIGIKTOGARQKPSLSYISQ 1206
Db 2041 GOYGOWAPKARLHYSGSINAWSTKEPFSWIKVDLAPMIHIGIKTOGARQKPSLSYISQ 2100
QY 1207 FIIMYSLDGKKWOTYRGNSTGLMVFNGVDSGIIKHNIENPPIIARIYIRLHPTHYSIRS 1266
Db 2101 FIIMYSLDGKKWOTYRGNSTGLMVFNGVDSGIIKHNIENPPIIARIYIRLHPTHYSIRS 2160
QY 1267 TLRMELMGCDLNSCSMPLGNEKSAISDAQITASSYFTNMFATWSPSKARLHLOGRSNAMR 1326
Db 2161 TLRMELMGCDLNSCSMPLGNEKSAISDAQITASSYFTNMFATWSPSKARLHLOGRSNAMR 2220
QY 1327 PQVNNPKWLOVDFQKTMKVTGVTQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKV 1386
Db 2221 PQVNNPKWLOVDFQKTMKVTGVTQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKV 2280
QY 1387 KVFQGNQDSTPVPVNSLDPPLRYLRHPQSVWQHIALRMEVLGCEAODLY 1438
Db 2281 KVFQGNQDSTPVPVNSLDPPLRYLRHPQSVWQHIALRMEVLGCEAODLY 2332

RESULT 8

US-10-133-907-4
; Sequence 4, Application US/10133907
; Publication No. US20030195223A1
; GENERAL INFORMATION:
; APPLICANT: Chien, Kenneth R
; APPLICANT: Hoshijima, Masahiko
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VII
; TITLE OF INVENTION: with vesicle vector
; FILE REFERENCE: 6627-PAL170
; CURRENT APPLICATION NUMBER: US/10/133,907
; PRIOR FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: 60/286,314
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 2351
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-133-907-4

Query Match 94.0%; Score 7227; DB 12; Length 2351;
Best Local Similarity 61.6%; Pred. No. 0;
Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;
QY 1 ATRRYLGVAVELSWDYNQSDLGELPVDARPPRPVKSPFFNTSVVYKTKLTFVEFTVHLEN 60
|||||

Db 20 ATRRYLGVAVELSWDYNQSDLGELPVDARPPRPVKSPFFNTSVVYKTKLTFVEFTVHLEN 79
QY 61 IAKPRPFWMLGLGPTTQAEVYDVTWITLKNMASHPVSLHAGVSYWKASEGAYDDQTSQ 120
Db 80 IAKPRPFWMLGLGPTTQAEVYDVTWITLKNMASHPVSLHAGVSYWKASEGAYDDQTSQ 139
QY 121 REKEDDKVPFGSGHTTVVQVLKENGPMASDPICLTYSYLSHVDLVKDLNSGLIGALLVCR 180
Db 140 REKEDDKVPFGSGHTTVVQVLKENGPMASDPICLTYSYLSHVDLVKDLNSGLIGALLVCR 199
QY 181 EGS LAKEKOTLHKFTLLFAVFDGKSWHSETKNSLMODRDAASAPAKPMHTVNGVNR 240
Db 200 EGS LAKEKOTLHKFTLLFAVFDGKSWHSETKNSLMODRDAASAPAKPMHTVNGVNR 259
QY 241 SLFGLIGCHRKSYVMHVIQMGTTPEVHSIFLEGTFLVRNHRQASLEISPIITFLTAQTLL 300
Db 260 SLFGLIGCHRKSYVMHVIQMGTTPEVHSIFLEGTFLVRNHRQASLEISPIITFLTAQTLL 319
QY 301 MDLGOPLLCHISSHOHDCWEATVKVDSCEPQLRMKNNEAEADYDDDLTDEMDVVRP 360
Db 320 MDLGOPLLCHISSHOHDCWEATVKVDSCEPQLRMKNNEAEADYDDDLTDEMDVVRP 379
QY 361 DDNSPSFQIRSVAKKHPKTVWHYIAAEEEDWDYAPLVADDDRYSKYQLNNGPQIG 420
Db 380 DDNSPSFQIRSVAKKHPKTVWHYIAAEEEDWDYAPLVADDDRYSKYQLNNGPQIG 439
QY 421 RYKVKYRFMAYTDETFKTRERAIQHESGILGPLLYGEVGTLLIIFKNQASRPNIYPHGI 480
Db 440 RYKVKYRFMAYTDETFKTRERAIQHESGILGPLLYGEVGTLLIIFKNQASRPNIYPHGI 499
QY 481 TDVRPLYSRRLPGVKVHLKDFPILPGEIFKYKWTVTVEGPTKSDPRCLTRYSSYVNM 540
Db 500 TDVRPLYSRRLPGVKVHLKDFPILPGEIFKYKWTVTVEGPTKSDPRCLTRYSSYVNM 559
QY 541 RDLASGLIGPLLIICYKESVDQRCNOIMSKENVILFSDENESWYLTENIQBFLENPAG 600
Db 560 RDLASGLIGPLLIICYKESVDQRCNOIMSKENVILFSDENESWYLTENIQBFLENPAG 619
QY 601 VOLEDEPFOASINMHSINGVYFDSQLSVCLHEVAYWYILSIGAQDTFLSVFSSGTFPKH 660
Db 620 VOLEDEPFOASINMHSINGVYFDSQLSVCLHEVAYWYILSIGAQDTFLSVFSSGTFPKH 679
QY 661 KMVYEDTLTLFPFSGETVFMENPGLWILGCHNSDFRNRGMTALLKVSXCDKNTGDY 720
Db 680 KMVYEDTLTLFPFSGETVFMENPGLWILGCHNSDFRNRGMTALLKVSXCDKNTGDY 739
QY 721 DSYEDISAYLLSKNNAIEPRSF 742
Db 740 DSYEDISAYLLSKNNAIEPRSF 799
QY 743 742
Db 800 KIQNVSSDILMLLROSPTPHGLSLSDLOEAKVETSDDPSPCAIDNNSLSNTHFRPQ 859
QY 743 742
Db 860 LHHSGDMVTPPSGLQLRLNEKLGTTAATLKLKDFKVSSTNNLSTIPSDMLAAGTDN 919
QY 743 742
Db 920 TSSLGPPSPMVHYDSOLDTTLFGKKSSPLTESGGPLSLSEENNDKSLSEGLMNSQESSW 979
QY 743 742
Db 980 GKNVSTESGRLFKGKRAHGAPALLTKDNALFKVSI LKTKNTKTSNNSATNRKTHIDGPSL 1039
QY 743 742
Db 1040 LIENSPVWQNIILSDTESFEKVKVTFPLIHDRMLMDKNATALRLNHSNKTSSKNWEMVQK 1099
QY 743 742
Db 1100 KEGPIPPDAQNDPMSFFKMLFLPESARWIQRTGKNSLNSGQGPSKQLVSLGPEKSV 1159

QY 743 ----- 742
Db 1160 QNFLSEKNNVVGKEFTKQVGLKEMVPPSSRNLFNLNHNHNTHNQEKLOEBIEK 1219
QY 743 ----- 742
Db 1220 KETLIQENVVLPOIHTVGTGKFMKULFLLSTRQNVESYDGAVAPVLQDFRSLNDSTNR 1279
QY 743 ----- 742
Db 1280 TKHTAHFSSKGBENLEGLGNQTKQIVEKYACTTRISPNTSQNFVTVQSRKALKOPRL 1339
QY 743 ----- 742
Db 1340 PLEETELEKRIIVDDTSTQNSKNMKHLTPSTLTQIDYNEKEGALTQSPSLDCLTRSHSI 1399
QY 743 ----- 742
Db 1400 PQANRSLPIAKVSFPSPIRPIYLTRVLFDQNSSHLPAASYRKQDSGVQBSHFLQAKK 1459
QY 743 ----- 742
Db 1460 NNLSAILTLEMTGDQREVSGISGATNSVTYKKVENTVLPKPLPXTSGKVLLPKVHI 1519
QY 743 ----- 742
Db 1520 YQKDLFPPTETNSGPHLDLVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESSAKTPSK 1579
QY 743 ----- 742
Db 1580 LLDPLANDHNYGTQIPKEBKWSQEKSPKTAFFKXDTTILSNACSNHAIAINBQONKP 1639
QY 743 ----- 742
Db 1640 EIEVTWAKQGRTERLCSQPPVLKQREBITRTTQSDQEEIDYDDTISVENKKEDFDIY 1699
QY 787 DDENQSPRSFQKTRHYFIAVERLDYGMSSPHVLNRASQSVPOFKVQFQEFID 846
Db 1700 DEDNQSPRSFQKTRHYFIAVERLDYGMSSPHVLNRASQSVPOFKVQFQEFID 1759
QY 847 GSFTQPLYRGELNEHLGLPGPIRAEVEDNIWTFRNQASRPYSFYSSLSIYEEORQGA 906
Db 1760 GSFTQPLYRGELNEHLGLPGPIRAEVEDNIWTFRNQASRPYSFYSSLSIYEEORQGA 1819
QY 907 EPRKNFVKPNETKTYFWKQVHMAPTKDBFCCKAWAYFSDVLEKDVHSGLIGPLLVCHT 966
Db 1820 EPRKNFVKPNETKTYFWKQVHMAPTKDBFCCKAWAYFSDVLEKDVHSGLIGPLLVCHT 1879
QY 967 NTLNPAHGRQVTVQEEALPFTTIEDTKSNVFTENMERNCRAPCNIOEMDPTFKENYRFA 1026
Db 1880 NTLNPAHGRQVTVQEEALPFTTIEDTKSNVFTENMERNCRAPCNIOEMDPTFKENYRFA 1939
QY 1027 INGIYMDTLPGLVMAQDQRIWYLLSNGSNENIHSIHFGSHVFTVRKBEYKVALYNLYP 1086
Db 1940 INGIYMDTLPGLVMAQDQRIWYLLSNGSNENIHSIHFGSHVFTVRKBEYKVALYNLYP 1999
QY 1087 GVFTVEMLPKAGIWRVECLIGEHLHAGMSTFLVYGNKQOTPLGMAASHIRDFQITAS 1146
Db 2000 GVFTVEMLPKAGIWRVECLIGEHLHAGMSTFLVYGNKQOTPLGMAASHIRDFQITAS 2059
QY 1147 GOYGOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIIHGIKTQABQKQSSLYISQ 1206
Db 2060 GOYGOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIIHGIKTQABQKQSSLYISQ 2119
QY 1207 FIWYSDGKKWQYTRNGSTGLMVFFGNVDSSGIKHNFNPPIIARYIRLHPHTYSIRS 1266
Db 2120 FIWYSDGKKWQYTRNGSTGLMVFFGNVDSSGIKHNFNPPIIARYIRLHPHTYSIRS 2179
QY 1267 TLRMELMGCDLNSCSMPLGMESKASIDAQITASSYFTNMFPATWSPSKARLHLQGRSNAR 1326
Db 2180 TLRMELMGCDLNSCSMPLGMESKASIDAQITASSYFTNMFPATWSPSKARLHLQGRSNAR 2239

QY 1327 PQVNNPKEMLOVDFQKTMKVGTGTTQGVKSLTSMYKBEFLISSQDQHWTLFFQNGKV 1386
Db 2240 PQVNNPKEMLOVDFQKTMKVGTGTTQGVKSLTSMYKBEFLISSQDQHWTLFFQNGKV 2299
QY 1387 KVFQGNQDSFTPVVNSLDPLLRYLRIHPQSWHQAIALRMEVLGCEAODLY 1438
Db 2300 KVFQGNQDSFTPVVNSLDPLLRYLRIHPQSWHQAIALRMEVLGCEAODLY 2351
RESULT 9
US-10-132-829-4
; Sequence 4, Application US/10132829
; Publication No. US2003004982A1
; GENERAL INFORMATION:
; APPLICANT: Chien, Kenneth R
; APPLICANT: Hoshijima, Masahiko
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor V
; TITLE OF INVENTION: with vesicle vector
; FILE REFERENCE: 6627-PALL70
; CURRENT APPLICATION NUMBER: US/10/132,829
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: 60/286,314
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-132-829-4
Query Match 94.0%; Score 7227; DB 15; Length 2351;
Best Local Similarity 61.6%; Pred. No. 0;
Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;
QY 1 ATRRYYLGAVELSWNDYMQSDIGELPVDARFPPRPVKSPFNTSVVYKKTFLFVEFTVHLFN 60
Db 20 ATRRYYLGAVELSWNDYMQSDIGELPVDARFPPRPVKSPFNTSVVYKKTFLFVEFTVHLFN 79
QY 61 TAKPRPPMGLLGPTIOAEVDTVTITLKNASHPVSLHVGVSYKASEGAEYDDQTSQ 120
Db 80 TAKPRPPMGLLGPTIOAEVDTVTITLKNASHPVSLHVGVSYKASEGAEYDDQTSQ 139
QY 121 REKEDDKVPFGSGSHYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
Db 140 REKEDDKVPFGSGSHYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 199
QY 181 EGSIAKEKOTLHKILLFAPVDEKSWHSETKNSLMQDRDAASARAMPKMTVNGYVNR 240
Db 200 EGSIAKEKOTLHKILLFAPVDEKSWHSETKNSLMQDRDAASARAMPKMTVNGYVNR 259
QY 241 SLPGILGCHRKSVYWHVIGMTTPEVHSITFLEGHTFLVNRHQASLEISPTIFLTAQTLL 300
Db 260 SLPGILGCHRKSVYWHVIGMTTPEVHSITFLEGHTFLVNRHQASLEISPTIFLTAQTLL 319
QY 301 MDLQGFLLFCHTSSHQHDMGEAYVVDSCPEPQLRMKNNEAEYDDDLTDSMDVVRP 360
Db 320 MDLQGFLLFCHTSSHQHDMGEAYVVDSCPEPQLRMKNNEAEYDDDLTDSMDVVRP 379
QY 361 DDNSPSPFQIRSAVKHGPKTWHYIAAEEEDWDYAPLVADDDRSYKSOYLNGPQRTG 420
Db 380 DDNSPSPFQIRSAVKHGPKTWHYIAAEEEDWDYAPLVADDDRSYKSOYLNGPQRTG 439
QY 421 RYKVKVRFWAYTDETFKTRREALQHSGLIGLPGLLYGEVGTLLIIFKNQASRPNTYPHGI 480
Db 440 RYKVKVRFWAYTDETFKTRREALQHSGLIGLPGLLYGEVGTLLIIFKNQASRPNTYPHGI 499
QY 481 TDVRLPYRRLPKGVKHLXDFPILPGEIFKYKMTVTVEDGPTKSDPRCLTRYSSFVNME 540
Db 500 TDVRLPYRRLPKGVKHLXDFPILPGEIFKYKMTVTVEDGPTKSDPRCLTRYSSFVNME 559
QY 541 RDLASGLIGLPGLLYCYKESVDQGNQIMSDKGNVILFVSFDENRSWTLTNIQRPFLNPAG 600

Db 560 RDLASGLIGLLICYKESVDQRGNQIMSDKRNVLFSVFDENRSWLTENIQRLPNPAG 619
Qy 601 VQLEDFEQASNMHSINGVFDLSQLSVCLHEVAYWYIISIGAQTDFLSVFFSGYTFKH 660
Db 620 VQLEDFEQASNMHSINGVFDLSQLSVCLHEVAYWYIISIGAQTDFLSVFFSGYTFKH 679
Qy 661 KMVEDTLTLPFSGGETVFMENPGLWILGCHNSDFRNRGWTALLKVSSCDKNTGDIYE 720
Db 680 KMVEDTLTLPFSGGETVFMENPGLWILGCHNSDFRNRGWTALLKVSSCDKNTGDIYE 739
Qy 721 DSYEDISAYLLSKNNAIEPRSF----- 742
Db 740 DSYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTIPENDIEKTDWFAHRTMP 799
Qy 743 ----- 742
Db 800 KIQNVSSDLLMLLRQSPTEPHGLSLDLQRAKYETFDSPGAGIDSNNSISEMTHFRPQ 859
Qy 743 ----- 742
Db 860 LHHSGDMVFTPEGLQLRLNEKLGTTAATBELKKLDFKVSSTNNLIISTIPSDNLAAGTDN 919
Qy 743 ----- 742
Db 920 TSSLGPPSPMPVHDSQLDITLPGKSSPLTESGGPLSLSEENNDKLLSGLMNSQESSW 979
Qy 743 ----- 742
Db 980 GKNVSSGESGLFKGKRAHPALLTKDNALFKVSI SLLKTNKTSNNSATNRKTHIDGPSL 1039
Qy 743 ----- 742
Db 1040 LIENSPSVWONILESDTEFKKVTPLIHDRMLMDKNATALRLNHNKNTSSKNMEMVQK 1099
Qy 743 ----- 742
Db 1100 KEGPIPPDAQNPDMSFFKMLFLPESARWI ORTHGKNSLNSGQSPKQLVSLGPEKSVEG 1159
Qy 743 ----- 742
Db 1160 QNFLSEKNKVVVGKEFTKDVGLXEMVPPSRNLFLTWLDNLHENNTNQEKKIQERIEK 1219
Qy 743 ----- 742
Db 1220 KETLIQENWVLPQIHVTGTRKFMKNLFLLLSTRQNVESYDGAAPVLQDPRSLDSTNR 1279
Qy 743 ----- 742
Db 1280 TKKHTAHFSKKGEENLEGLGNQTKQIVEKYACTTRISPNTSQONFVTRSKRALKQFRL 1339
Qy 743 ----- 742
Db 1340 PLEETELEKRIIVDDTSTQSKNMKHLTPSLTQIDYNEKEGALTQSPSLDCLTRSHSI 1399
Qy 743 ----- 742
Db 1400 PQANRSLPIAKVSSFSIRPIYLRVLFDNSSHLPAAVSRKXDSGVQESSHFLQAKK 1459
Qy 743 ----- 742
Db 1460 NNLSLAILTLEMTGDQREVGLGTSATNSVTYKKVENTVLPKPLPKTSKGKVELLPKVI 1519
Qy 743 ----- 742
Db 1520 YQKDLFPPTETSGSPGHLDLVEGSLLOQTEGAIKWNEARPKVPFLRVATESAKTPSK 1579
Qy 743 ----- 742
Db 1580 LLDPLAWDNHYGTQIPKEWKSQEKSPKTAPEKKKDTILSLNACESNHAAIAINEGQMKP 1639
Qy 743 -----SONPVLKXHQREIRTRTLQSDQBEIDYDDTISVEMKKEDFDIY 786
Db 1640 EIEVTWAKQGRTERLCSQNPVLRHQREIRTRTLQSDQBEIDYDDTISVEMKKEDFDIY 1699

Qy 787 DEENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHYLRNRAQSGSVPOKKVVFQBFETD 846
Db 1700 DEENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHYLRNRAQSGSVPOKKVVFQBFETD 1759
Qy 847 GSTQPLYRGELNEHLLGLGPYIRAEVDNIMVTERNOASRPYSFYSSLSIYEDORQGA 906
Db 1760 GSTQPLYRGELNEHLLGLGPYIRAEVDNIMVTERNOASRPYSFYSSLSIYEDORQGA 1819
Qy 907 EPRKNFVKNETKTYFMKVQHMAPTKDEPDCXAMAYFSDVDLEKDVHSGLIGPLVCHT 966
Db 1820 EPRKNFVKNETKTYFMKVQHMAPTKDEPDCXAMAYFSDVDLEKDVHSGLIGPLVCHT 1879
Qy 967 NTLNPAHGRQVTVQEFALPFTIIFDETKSWYFTFMERNCRAPCNIOHEDPTFKENTRPHA 1026
Db 1880 NTLNPAHGRQVTVQEFALPFTIIFDETKSWYFTFMERNCRAPCNIOHEDPTFKENTRPHA 1939
Qy 1027 INGVIMDTLPLGLYMAQDQIRWYLLSGNSNENHSHFSGHVTFRKCKRYKVALYNLYP 1086
Db 1940 INGVIMDTLPLGLYMAQDQIRWYLLSGNSNENHSHFSGHVTFRKCKRYKVALYNLYP 1999
Qy 1087 GFVETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPPLGWSGHIRDFQITAS 1146
Db 2000 GFVETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPPLGWSGHIRDFQITAS 2059
Qy 1147 GOYGWAPKARLHYSGSINAWSTKBPFSWIKVDLLAPMIIHGIKTOGARQKFSLSYSQ 1206
Db 2060 GOYGWAPKARLHYSGSINAWSTKBPFSWIKVDLLAPMIIHGIKTOGARQKFSLSYSQ 2119
Qy 1207 FIIMYSIDGKKWOTYRGNSGTGLWVFGNVDSGSIKHNIENPPIIARIYIRLHPHYSIRS 1266
Db 2120 FIIMYSIDGKKWOTYRGNSGTGLWVFGNVDSGSIKHNIENPPIIARIYIRLHPHYSIRS 2179
Qy 1267 TLRMELMGCDLNSCSPGLMGESKAI SDAQITASSYFTNMFPATWSPSKARLHLOQRSNAWR 1326
Db 2180 TLRMELMGCDLNSCSPGLMGESKAI SDAQITASSYFTNMFPATWSPSKARLHLOQRSNAWR 2239
Qy 1327 PQVNNPKEMLOVDPOKTMKVTVGTQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKV 1386
Db 2240 PQVNNPKEMLOVDPOKTMKVTVGTQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKV 2299
Qy 1387 KVFQGNQDSTPVVNSLDPELLTRYLRHPQSWVHQIATRMVGLGCEAODLY 1438
Db 2300 KVFQGNQDSTPVVNSLDPELLTRYLRHPQSWVHQIATRMVGLGCEAODLY 2351
RESULT 10
US-10-172-712-27
; Sequence 27, Application US/10172712
; Publication No. US20030125232A1
; GENERAL INFORMATION:
; APPLICANT: GRIFFIN, JOHN H.
; APPLICANT: GALE, ANDREW J.
; APPLICANT: GERTZOFF, ELIZABETH D.
; APPLICANT: PELLEGUER, JEAN-LUC
; TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS
; FILE REFERENCE: 4198-4001US1
; CURRENT APPLICATION NUMBER: US/10172,712
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: 60/238,578
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-172-712-27

Query Match 94.0%; Score 7227; DB 15; Length 2351;
Best Local Similarity 61.6%; Pred. No. 0;
Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;

QY	1	ATRYVYLGAVALSMDYMSDGLGELPVDARPPRPVPSKPPFNTSVVYKXKTLFVBTZVHLFN	60
Db	20	ATRYVYLGAVALSMDYMSDGLGELPVDARPPRPVPSKPPFNTSVVYKXKTLFVBTZVHLFN	79
QY	61	IAKPRPPMGLLGPTIQAEVYDVTVITLKNMASHPVSLHVGVSYWKASGAEYDDQTSQ	120
Db	80	IAKPRPPMGLLGPTIQAEVYDVTVITLKNMASHPVSLHVGVSYWKASGAEYDDQTSQ	139
QY	121	REKEDKVPFGSHYVQVQLKENGPMASDPLCLTYSYLSHVVDLYKDNLSGLIGALLVCR	180
Db	140	REKEDKVPFGSHYVQVQLKENGPMASDPLCLTYSYLSHVVDLYKDNLSGLIGALLVCR	199
QY	181	EGSLAKEKTOTLHKFILLPAVDEGKSWHSETKNSLMODRDAASARAMPKMHYVNGYNR	240
Db	200	EGSLAKEKTOTLHKFILLPAVDEGKSWHSETKNSLMODRDAASARAMPKMHYVNGYNR	259
QY	241	SLPGLIGCHRSKVYWHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTL	300
Db	260	SLPGLIGCHRSKVYWHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTL	319
QY	301	MDLGQFLLPCHTSSHOHDGMEAYVVDSCPEPQPMKNNEAEYDDDLTDSMDVVRP	360
Db	320	MDLGQFLLPCHTSSHOHDGMEAYVVDSCPEPQPMKNNEAEYDDDLTDSMDVVRP	379
QY	361	DDNSPSFTQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLPDDRYSQYLLNNGPQRIG	420
Db	380	DDNSPSFTQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLPDDRYSQYLLNNGPQRIG	439
QY	421	RKYKVRFMAYTDTFTKTRAIQHESGILGPLLYGEVGTLLIIIFKNQASRYNIYPHGI	480
Db	440	RKYKVRFMAYTDTFTKTRAIQHESGILGPLLYGEVGTLLIIIFKNQASRYNIYPHGI	499
QY	481	TDVRPLYRRLPKGVKHLKDFPILPCEIFKYKWTVTVEGPTKSDPRCLTRYSSVFNME	540
Db	500	TDVRPLYRRLPKGVKHLKDFPILPCEIFKYKWTVTVEGPTKSDPRCLTRYSSVFNME	559
QY	541	RLASGLIGPLLCYKESVDQRGNQJMSDKRNVILFSVFDENRSWYLTENIQRFLENPAG	600
Db	560	RLASGLIGPLLCYKESVDQRGNQJMSDKRNVILFSVFDENRSWYLTENIQRFLENPAG	619
QY	601	VOLEDEFOASNMHISINGVVDLSQLSVCLHEVAYWYILSIGAQTDFLSVFPSTFKH	660
Db	620	VOLEDEFOASNMHISINGVVDLSQLSVCLHEVAYWYILSIGAQTDFLSVFPSTFKH	679
QY	661	KMYVEDTLPLPFGSETVFMENPGLWILGCHNSDFRNGMTALLKVSQCDQNTGDYIE	720
Db	680	KMYVEDTLPLPFGSETVFMENPGLWILGCHNSDFRNGMTALLKVSQCDQNTGDYIE	739
QY	721	DSYEDISAYLLSKNNAIEPRSF-----	742
Db	740	DSYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATIPENDIEKTDPPFAHRTMP	799
QY	743	-----	742
Db	800	KIQNVSSDILLMLLRQSPHGLSLDLQEAKEYTFSDPSPCAIDSNNSLSBMTFRPQ	859
QY	743	-----	742
Db	860	LHSGDMVFTPEGLQRLINEXLGTAAATELKKLDPKVSSTSNLITIPSDNLACTDN	919
QY	743	-----	742
Db	920	TSSLGPPMPVHYSQDQTLTLFGKSSPLTESGGPLSBEENDSKLLSGLMNSQESSW	979
QY	743	-----	742
Db	980	GKNVSTESGRLFKGKRAHGPAALLTDNALFKVSIKLTNKTNSNATNRKTHIDGPSL	1039
QY	743	-----	742
Db	1040	LIENSPSVWQNIESTDEFKKVTLIHDRMLMDKNATRLNHNKNTSSKNMVMQOK	1099
QY	743	-----	742

Db	1100	KEGIPPPDAQNPMDSFFKMLFLPESARWIORTHGKNSLNSGQSPSKQLYSLGPEKSVBG	1159
QY	743	-----	742
Db	1160	QNFLEKKNVVGKEFTKDVGLKEMVFPSSRNLFNLNLDNLHNNTHNQEKIQBIEIK	1219
QY	743	-----	742
Db	1220	KETLIQENVVLPQIHVTGTGKNFMKNLFLSTRQNVGSDYGAAYVLQDFRSLDSTNR	1279
QY	743	-----	742
Db	1280	TKKHTAHFSKKGBENLEGLNQTKQVEKVIACVTRISPNTSQNFVQTSRKALKQFRL	1339
QY	743	-----	742
Db	1340	PLEBTELEKRIIVDDTSTQMSKNMKHLTPSTLTQIDYNEKEKGALITQSPSLDCLTRSHI	1399
QY	743	-----	742
Db	1400	POANRSLPIAKVSFSPSIRPIYLTVLFQDNSSHLPAASYRKDQSGVQESSHFLQGAKK	1459
QY	743	-----	742
Db	1460	NNLSAILTLEMTGDQREVGLGTSATNSVYKKVENTVLPKPLPKTSGKVELLPKVIH	1519
QY	743	-----	742
Db	1520	YQKDLFPPTETSNGLDVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESSAKTPSK	1579
QY	743	-----	742
Db	1580	LILDPLAWNHYGTQIPKEBWSQEKSPKTAFFKDDTILSLNACSNHAIAINSGONKP	1639
QY	743	-----SONPPVLKEHOREITRTTLQSDQEBTDYDDTISVEMKXEDFDIY	786
Db	1640	EIEVTWAKQGRTERLCSQNPVLRKHOREITRTTLQSDQEBTDYDDTISVEMKXEDFDIY	1699
QY	787	DDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRAQSGSVPOFKVVFQEFYD	846
Db	1700	DDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRAQSGSVPOFKVVFQEFYD	1759
QY	847	GSFTQPLXGELENEHGLLGPYIRAEVDNIIMVTRNQASRPYSFVSSLSIYEEBQOQA	906
Db	1760	GSFTQPLXGELENEHGLLGPYIRAEVDNIIMVTRNQASRPYSFVSSLSIYEEBQOQA	1819
QY	907	EPKRFVKNETKTYFWKQHEMAPTKDQDFCKAWAYPSDVKDQVHSGLLIGPLIVCHT	966
Db	1820	EPKRFVKNETKTYFWKQHEMAPTKDQDFCKAWAYPSDVKDQVHSGLLIGPLIVCHT	1879
QY	967	NTLNPAGRQVTVQEPALPFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTPKENYRFA	1026
Db	1880	NTLNPAGRQVTVQEPALPFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTPKENYRFA	1939
QY	1027	INGYIMDTLPLGLVMAQDQIRWYLLSNGNENIHSIHFSGHVPVTKKXEYKXALYNLYP	1086
Db	1940	INGYIMDTLPLGLVMAQDQIRWYLLSNGNENIHSIHFSGHVPVTKKXEYKXALYNLYP	1999
QY	1087	GUPETVEMLPKAGIWRVCLIGEHLHAGMSTLFLVYNNKOTPLGMAASHIRDFQITAS	1146
Db	2000	GUPETVEMLPKAGIWRVCLIGEHLHAGMSTLFLVYNNKOTPLGMAASHIRDFQITAS	2059
QY	1147	GOYGOWAPKLARLHYSGSINAMSTKEPFSWIKVYDLPAMIHIGKITQGAQKQFSSYISQ	1206
Db	2060	GOYGOWAPKLARLHYSGSINAMSTKEPFSWIKVYDLPAMIHIGKITQGAQKQFSSYISQ	2119
QY	1207	FIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSGIGKHNIPNPPIIARIYRLHPTHYSIRS	1266
Db	2120	FIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSGIGKHNIPNPPIIARIYRLHPTHYSIRS	2179
QY	1267	TLRMELMGCDLNSCSPLGMSKXVIAISDAQITASSYFTNPEATWSPSKARLHLOGRNAR	1326

TITLE OF INVENTION: Modified Factor VIII
FILE REFERENCE: 75-95J
CURRENT APPLICATION NUMBER: US/10/131,510A
CURRENT FILING DATE: 2002-07-10
PRIOR APPLICATION NUMBER: U.S. 09/315,179
PRIOR FILING DATE: 1999-05-20
PRIOR APPLICATION NUMBER: U.S. 09/037,601
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: U.S. 08/670,707
PRIOR FILING DATE: 1996-06-26
PRIOR APPLICATION NUMBER: PCT/US97/11155
PRIOR FILING DATE: 1997-06-26
PRIOR APPLICATION NUMBER: PCT/US94/13200
PRIOR FILING DATE: 1994-11-15
PRIOR APPLICATION NUMBER: U.S. 08/212,133
PRIOR FILING DATE: 1994-03-11
PRIOR APPLICATION NUMBER: U.S. 07/864,004
PRIOR FILING DATE: 1992-04-07
NUMBER OF SEQ ID NOS: 40
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 39
LENGTH: 1443
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: factor VIII lacking
OTHER INFORMATION: the B domain
US-10-131-510A-39

Query Match 84.4%; Score 6489; DB 12; Length 1443;
Best Local Similarity 83.4%; Pred. No. 0;
Matches 1200; Conservative 104; Mismatches 119; Indels 16; Gaps 3;

1 ATRRYLGAVELSWDYMQSD-LGRLPVDARFPKSPFTSVVYKTLFVETVHLF 59
20 AIRRYLGAVELSWDYMQSD-LGRLPVDARFPKSPFTSVVYKTLFVETVHLF 79
60 NTAERPPMGLGPTIQAEVYDVTITLKNASHPVSLHAGVSVKASEGAEYDDQTS 119
80 SVARPPMGLGPTIQAEVYDVTITLKNASHPVSLHAGVSVKASEGAEYDDQTS 139
120 QREKEDKVPFGSSHTYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVC 179
140 QREKEDKVPFGSSHTYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVC 199
180 REGSLAKEKVTQHLKFIILFAVDECKSWHSETKNSLMQDRDAASARAPKHTVNGYN 239
200 REGSLAKEKVTQHLKFIILFAVDECKSWHSETKNSLMQDRDAASARAPKHTVNGYN 259
240 RSLPGLIGCHRSVYWHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL 299
260 RSLPGLIGCHRSVYWHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL 319
300 LMDLQGLFLFCHISSHHGMEARVSVCAEPPQLRKNNEAEYDDDLTSEMDVVR 359
320 LMDLQGLFLFCHISSHHGMEARVSVCAEPPQLRKNNEAEYDDDLTSEMDVVR 378
360 FDDNDSFPTQIRSVAKKPKTWVHYIAAEERDWDYAPIVLAPDPSYKSOYLNNGPQRI 419
379 LDGDDVSPFTQIRSVAKKPKTWVHYIAAEERDWDYAPIVLAPDPSYKSOYLNNGPQRI 438
420 GRVYKVRFWAYTDEPKTEATQHEGILGPLYGEVGTDLIIIFKQASRPYNYPHG 479
439 GRVYKVRFWAYTDEPKTEATQHEGILGPLYGEVGTDLIIIFKQASRPYNYPHG 498
480 ITDVRPLYSRLPRGVKHLKDFILPGEIFKYKWTVTVEGPTKSDPRCLTRYSSFVNM 539
499 ITDVRPLYSRLPRGVKHLKDFILPGEIFKYKWTVTVEGPTKSDPRCLTRYSSFVNM 558
540 ERDLASGLIGPLIICYKESVDQGNQIMSKRNVILFSDENRPSWILTENIQRFLNPA 599
559 EKDLASGLIGPLIICYKESVDQGNQIMSKRNVILFSDENRPSWILTENIQRFLNPD 618

600 GVQLDEDFQASNMHSINGVYFDSLOLSVCLHEVAYVWYILSIGAOTDFLSVPFSGVTFK 659
619 GLPQDFQASNMHSINGVYFDSLOLSVCLHEVAYVWYILSIGAOTDFLSVPFSGVTFK 678
660 HKWYEDTLLTLPFSGETVFMSENPGILWILGCHNSDFRNGMTALLKVSCKNTGDYY 719
679 HKWYEDTLLTLPFSGETVFMSENPGILWILGCHNSDFRNGMTALLKVSCKNTGDYY 738
720 EDSYEDISAYLLSKNAIEPRSPQNPVPLKQREITRTTLQSDQBEIDYDDTISVEMK 779
739 DNTYEDIPGFLLSGKNIEP-----RDISLFTQPEEDKMDYDDIFSTETK 784
780 KEDFDIYDEENQSPRSFQKTRHYFTAAVERLDWYGMSSSPHYLRRAQSGSPQPKV 839
785 GEFDIYDEENQSPRSFQKTRHYFTAAVERLDWYGMSSSPHYLRRAQSGSPQPKV 844
840 VFQSFDTGSGFTQPLRYGELNEHLGLLGPYIRAEVEDNIMVTFRQAASRPYSFYSSLISYE 899
845 VFREFADGSGFTQPLRYGELNEHLGLLGPYIRAEVEDNIMVTFRQAASRPYSFYSSLISYP 904
900 EDQSGAEPKRNFKVKNETKTYFWKQHHMAPTKDEDFCKAWAYPSVDLEKDVHSGLIG 959
905 DDQSGAEPKRNFKVKNETKTYFWKQHHMAPTKDEDFCKAWAYPSVDLEKDVHSGLIG 964
960 PLLVCHTWTINPAHGRQVTVQEFALFTTIDETKSWYFTENWERNCRAPCNIOMEDPTFK 1019
965 PLLICRANTLNAHGRQVTVQEFALFTTIDETKSWYFTENWERNCRAPCNIOMEDPTFK 1024
1020 ENYFHAINGYIMDTLPGLVMAQQRIRWYLLSGNSNENIHSIHFSGHVFTVRKKEEYKM 1079
1025 ENYFHAINGYIMDTLPGLVMAQQRIRWYLLSGNSNENIHSIHFSGHVFTVRKKEEYKM 1084
1080 ALYNYLPGVFTVEMLPSKAGIWRVECLIGHLHAGMTLFLVYSNKCQTPGLMASGHIR 1139
1085 AVYNYLPGVFTVEMLPSKAGIWRVECLIGHLHAGMTLFLVYSNKCQTPGLMASGHIR 1144
1140 DFOITASQYQGWAPKPLARLHYSGINAWSTKEPSPMLKVDLLAPMIHGLTKQCARQKF 1199
1145 DFOITASQYQGWAPKPLARLHYSGINAWSTKEPSPMLKVDLLAPMIHGLTKQCARQKF 1204
1200 SSLYISQIFIMYSLDGRWQSVRGNSTGTLMVFFGNVDSSGKKNIFNPPIIYIRLHP 1259
1205 SSLYISQIFIMYSLDGRWQSVRGNSTGTLMVFFGNVDSSGKKNIFNPPIIYIRLHP 1264
1260 THYSIRSTLRMLGCDLNSCMPLGMQNKATSDSQTITASSHLSNIFATWSPSQARLHLQ 1319
1265 THYSIRSTLRMLGCDLNSCMPLGMQNKATSDSQTITASSHLSNIFATWSPSQARLHLQ 1324
1320 GRSNARQVNNPKEMLOVDFOKTKVGTGVTQGVKSLTSMYVKEFLSSQDGHOWTL 1379
1325 GRSNARQVNNPKEMLOVDFOKTKVGTGVTQGVKSLTSMYVKEFLSSQDGHOWTL 1384
1380 PFQNGKVKVQGNQSDFTPVVNSLDPPLIYLRHPQSWHVIQIALRMEVLGCEAODLY 1438
1385 FLQDGHKVKVQGNQSDFTPVVNSLDPPLIYLRHPQSWHVIQIALRMEVLGCEAODLY 1443

RESULT 13
US-10-187-319-39
Sequence 39, Application US/10187319
Publication No. US20030068785A1
GENERAL INFORMATION:
APPLICANT: Lollar, John S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303
COMPUTER READABLE FORM:


```

;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/187,319
; FILING DATE: 27-Aug-2002
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/523,656
; FILING DATE: 2000-03-10
; APPLICATION NUMBER: US 09/037,601
; FILING DATE: 1998-03-10
; APPLICATION NUMBER: WO PCT/US97/11155
; FILING DATE: 1997-06-26
; APPLICATION NUMBER: US 08/670,707
; FILING DATE: 1996-06-26
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenlee, Loranace L.
; REGISTRATION NUMBER: 27,894
; REFERENCE/DOCKET NUMBER: 75-95K
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
;
; INFORMATION FOR SEQ ID NO: 39:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1443 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-10-187-319-39

Query Match      84.4%; Score 6489; DB 15; Length 1443;
Best Local Similarity 83.4%; Pred. No. 0;
Matches 1200; Conservative 104; Mismatches 119; Indels 16; Gaps 3

Qy      1  ATRYVYLGAVELSDYMQSD-LGELPVDARPPRPVKSPFNTSVVYKTLFVETVHLF 59
Db      20  AIRRYVYLGAVELSDYRQSELRLHVDTRFPATAPGALPLGPSVLYKTVFVEFTDQLF 79

Qy      60  NIAKPRPPWMLGPTTQAEVYDVTWITLKNWASHPVSLHVGVSYWKASGAEYDDOTS 119
Db      80  SVAPRPPWMLGPTTQAEVYDVTWITLKNWASHPVSLHVGVSFWKSSGAEYEDHTS 139

Qy      120  QREKEDDKVFPGGSHTVYQVLKNGPNMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVC 179
Db      140  QREKEDDKVLPGKSQTVYQVLKNGPTASDPCLTYSYLSHVDLVKDLNSGLIGALLVC 199

Qy      180  REGSLAKEKTOHLKHFILLPAFVDEGKSWHSETKNSLMQDRDAASARAPKQHTVNGYN 239
Db      200  REGSLTRERTQNLHEFVLLPAFVDEGKSWHGBANDSWTRAMDPAAPARQAPMHTVNGYN 259

Qy      240  RSLPGLIGCHRKSVYVHVGTTTPEVHSIFLEHTFLVRNHRQASLSIPTFLTAQTL 299
Db      260  RSLPGLIGCHKKSYYVHVGMTSPSEVHSIFLEHTFLVRHHRQASLSISPLTFLTAQTF 319

Qy      300  LMDLGQFLFCHISSHQHDGMEAYYKVDSCPEEPQLRMKNNEAEADYDDLTDSEMDVVR 359
Db      320  LMDLGQFLFCHISSHRRHGMEAHVRVSCAPEEPQLRREKADEE-EDYDNLNLYDSMDVVR 378

Qy      360  FDDNDSPSFIQIRSVAKKHPKTHVHYIAAREEDWDYAPLVLAAPPDRSKYSQYLNNGPQRI 419
Db      379  LDGDDVSPFIQIRSVAKKHPKTHVHYISAEEDWDYAFAPSPSPRSYKSLYNSGPQRI 438

Qy      420  GRKYKVRFMAYTDTFTKTRAIQIHESGLIGPLLYGEGVDTLLIIFKQASRPYNIYPHG 479
Db      439  GRKYKARFVAYTDVTFKTRAIPEYSGILGFLLYGEGVDTLLIIFKNKASRPYNIYPHG 498

Qy      480  ITDVRPLYSRRLPKGVKHLKQDPIILPGEIFKYKWTVTVEDGPKSDPCLCTRYYSFVM 539
Db      499  ITDVSALHPGRLKQWHLKQDPIILPGETFKYKWTVTVEDGPKSDPCLCTRYYSINL 558

```

;; PRIOR APPLICATION NUMBER: U.S. 09/037,601
;; PRIOR FILING DATE: 1998-03-10
;; PRIOR APPLICATION NUMBER: U.S. 08/670,707
;; PRIOR FILING DATE: 1996-06-26
;; PRIOR APPLICATION NUMBER: PCT/US97/11155
;; PRIOR FILING DATE: 1997-06-26
;; PRIOR APPLICATION NUMBER: PCT/US94/13200
;; PRIOR FILING DATE: 1994-11-15
;; PRIOR APPLICATION NUMBER: U.S. 08/212,133
;; PRIOR FILING DATE: 1994-03-11
;; PRIOR APPLICATION NUMBER: U.S. 07/864,004
;; PRIOR FILING DATE: 1992-04-07
;; NUMBER OF SEQ ID NOS: 40
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 37
;; LENGTH: 2133
;; TYPE: PRT
;; ORGANISM: Sus scrofa
US-10-131-510A-37

Query Match 81.0%; Score 6227; DB 12; Length 2133;

Best Local Similarity 57.3%; Pred. No. 0;
Matches 1212; Conservative 106; Mismatches 119; Indels 678; Gaps 3;

QY 1 ATRRYLGAVELSDWIMQSD-IGELFVDARPPRPVKSPFNSTVYKTLFVETVHLF 59
DB 20 AIRRYLGAVELSDWYRQSELLRELHVDTRFATAPGALPLGPSVLYKKTVPEVTDOLF 79
QY 60 NIAKRRPPMGLGPTIOAEVYDVTIVILKNASHPVSLHAGVSNWKASGAEVDDOTS 119
DB 80 SVAREPPMGLGPTIOAEVYDVTIVILKNASHPVSLHAGVSNWKASGAEVDDOTS 139
QY 120 QREKEDDKVPFGSGSHYVWOLKENGPMASDPLCLITYSLSHVDLVKDLNSGLIGALLVC 179
DB 140 QREKEDDKVLPKGSQTYVWQVLKENGPTASDPPCLITYSLSHVDLVKDLNSGLIGALLVC 199
QY 180 REGSLAKEKTQTLHFKFILLFAVDFDEKSWHSETKNSLMQDRDAASARAPKMTVNGYN 239
DB 200 REGSLTRETONLHERVLLFAVDFDEKSWHSEARNDSWTRAMPAPARAQPAHMTVNGYN 259
QY 240 RSLPLGIGCHRKSVYWHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTL 299
DB 260 RSLPLGIGCHKSVYWHVIGMGTSPVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTF 319
QY 300 LMDLGQFLFCHLSSHQHGMEYKVKVDCSPREPQLRMKNREAEYDDDLTDSMDVVR 359
DB 320 LMDLGQFLFCHLSSHQHGMEYKVKVDCSPREPQLRMKNREAEYDDDLTDSMDVVR 378
QY 360 FDDNSPSTQIRSVAKGHPKTVVHYIAAEEEDWDYAPLVAPDDRSYKSOYLNNQPORI 419
DB 379 LDGDDVSPFIQIRSVAKGHPKTVVHYISAEEDWDYAPVPSPDSRSYKSLYLNSGPORI 438
QY 420 GRKYKVRWYATDETFKTRALQHSGLIGLPGLYGEVGTLLIIFKNOASRPYNYLPHG 479
DB 439 GRKYKARFVAYDVTFKTRKALPYESGILGPLYGEVGTLLIIFKNOASRPYNYLPHG 498
QY 480 ITDVRPLYSRLPKGVKHLKDFPILGELFKYKWTVVEDGPKSPRCLTRYSSFNVM 539
DB 499 ITDVSALHPGRLLKWLKMLKMPILGELFKYKWTVVEDGPKSPRCLTRYSSFNVM 558
QY 540 ERDLASGLIGPLLIICYKESVDQRGNQIMSDKRVNLLFSVFDENRSWYLTENIORFLPNPA 599
DB 559 EKDLASGLIGPLLIICYKESVDQRGNQIMSDKRVNLLFSVFDENRSWYLTENIORFLPNPD 618
QY 600 GVQLEDEPQASNMHSINGVYFDSLOLSVCLHEVAYWYILSTGAQTDPLSVFSGYTEK 659
DB 619 GLQPDQEPQASNMHSINGVYFDSLOLSVCLHEVAYWYILSTGAQTDPLSVFSGYTEK 678
QY 660 HKWVYEDTLTLPFSGGTVMENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYY 719
DB 679 HKWVYEDTLTLPFSGGTVMENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYY 738
QY 720 EDSYEDISAYLLSKNAIEPRFSQN----- 745

Db 739 DNTYEDIPGFLSGKNVIEPRSPAQRPPPSASQKQPTTTSDEDDVELDPQSGERTQAL 798
QY 746 ----- 745
Db 799 EELSVPGDGSMLLGQNPAPHGSSSSDLQEARNEADDYLPQARERTAPSAARLRDELH 858
QY 746 ----- 745
Db 859 HSAERVLTPEPEKELKKLDSKMSSSDLLKTSPTIPSDTILSAETERTHSLGPPHPQVNR 918
QY 746 ----- 745
Db 919 SOLGAILVGNSSHFICAGVPLGSTEEDHESSLGENVSPVESDGIPEKERAHGASLTOKD 978
QY 746 ----- 745
Db 979 DVLFKVNSILVKTNKARVYLKTNKRIHIDDAALLTENRASATFMDKNTTASGLNHVSNWI 1038
QY 746 ----- 745
Db 1039 KGPLGNPLSSRGSPPELLTSSGSGKSVKQSSGQGRIRVAVEEBELSKGEMMLPNSE 1098
QY 746 ----- 745
Db 1099 LTFLTNSADVQNDTHSGGKKSREMERREKLVOEKVDLPQVYATGTGNFLRNIFHOST 1158
QY 746 ----- 745
Db 1159 EPSVEGPDGSHAPVPQDSRSLNDSABRAETHIAHFAISAREEAPLEAPNGRTGPPRSVA 1218
QY 746 ----- 745
Db 1219 PRRVKQSLKQIRLPLEEIKPERGVVLTATSTRWSESSPILOGAKNNLSLPLTLEMAGG 1278
QY 746 ----- 745
Db 1279 QOKISALGSAAGPLASGLKLEKAVLSSAGLSEASGAEFLPKVRVHREDLLPQKTSNVSC 1338
QY 746 ----- 745
Db 1339 AHGDLGQELFLQKTGPNVNLKNVNRPGRTPTSKLLGPPMPKEWSELESKPSKSTALRTKDI 1398
QY 746 ----- 745
Db 1399 SLPLDRHESNHSIAAKNEGQAEQTEAAATWQKGGPGLCAPKPPVLRHRQDISLPTQP 1458
QY 764 DQEEIDYDDTISVENMKKEDFDIYDEENQSPRSFQKTRHYFIAAVERLDYGMSSSPHV 823
Db 1459 EEDKMDYDDIIFSTETKGEDFDIYGEDENQDPSFQKTRHYFIAAVERLDYGMSSSPRA 1518
QY 824 LENRAQSGSVPOFKKVFQEFQDGSFTQPLYRGEINELHGLLGPYIRAEVEDNIMVTFN 883
Db 1519 LENRAQNEVEPRFKVVFREPADGFTQPSYRGEINELHGLLGPYIRAEVEDNIMVTFN 1578
QY 884 QASRPYSFVSSLSIIVEEDORQCAEPRKPFVKNEKTYFKWQOHHMAFKDPDCKAWAY 943
Db 1579 QASRPYSFVSSLSIIPDDQEQGAEPHNFVQNETRTYFKWQOHHMAFKDPDCKAWAY 1638
QY 944 FSDVLEKDVHSLGILLVCHTNTLNPAHQGVTVQEPALFTTIFDETKSWYFENMER 1003
Db 1639 FSDVLEKDVHSLGILLVCHTNTLNPAHQGVTVQEPALFTTIFDETKSWYFENMER 1698
QY 1004 NCRAPCNIOEDPTFKENYRFAHNGYIMDTLPLGLVMAQDQIRWYLLSMGSMENIHSIH 1063
Db 1699 NCRAPCHLQEDPTLKENYRFAHNGYIMDTLPLGLVMAQDQIRWYLLSMGSMENIHSIH 1758
QY 1064 FSGHVFTRKKEKYMALINLPGVFPETVEMLPSSKAGIWRVBECLTGEHLHAGMSTFLVY 1123
Db 1759 FSGHVFTRKKEKYMALINLPGVFPETVEMLPSSKAGIWRVBECLTGEHLHAGMSTFLVY 1818
QY 1124 SNKQOTPLGMSCHTEDFOITASGOVQWAPKLARLHYSGSINAMSTKEPPSWIKVILLIA 1183

Db 1819 SKECAPLGMASGRIRDFQITASGOYQOWAPKLARLHYSGSINAWSTDPHSWIKVDLLA 1878
 Qy 1184 PMIHGINTQARQKFSLLYSIQFIIMYSLDGKQWQTYRGNSGTGLMVFPGNVDSGGIKH 1243
 Db 1879 PMIHGINTQARQKFSLLYSIQFIIMYSLDGRNQSYRGNSGTGLMVFPGNVDSGGIKH 1938
 Qy 1244 NIFNPPITARYIRLPHYTHYSIRSTRMELMCGDLNCSMPILGMSKALSDAQITASSYFT 1303
 Db 1939 NIFNPPITARYIRLPHYTHYSIRSTRMELMCGDLNCSMPILGMSKALSDAQITASSHLS 1998
 Qy 1304 NWFATWSPSKARLHLOGRSNAWRPOVNNPKEMQLQVDFQTKMKTGVTGTTQGVKSLLTSMYV 1363
 Db 1999 NIFATWSQARLHLOGRSNAWRPRVSSAEWLQVDFQTKMKTGVTGTTQGVKSLLTSMYV 2058
 Qy 1364 KEFLISSQDGHQWTLFPQNGKVKVFCQGNQDSFTFPVNSLPPPLTRYLRIRHPOSWHQI 1423
 Db 2059 KEFLVSSQDGRWTLFLQDGHKTVFCQGNQDSSTEVVNALDPLPLTRYLRIRHPTSWAHI 2118
 Qy 1424 ALRMEVLGCEAQDLY 1438
 Db 2119 ALRMEVLGCEAQDLY 2133

RESULT 15

US-10-187-319-37
 ; Sequence 37, Application US/10187319
 ; Publication No. US20030068785A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lollar, John S.
 ; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
 ; NUMBER OF SEQUENCES: 40
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
 ; STREET: 5370 Manhattan Circle Suite 201
 ; CITY: Boulder
 ; STATE: Colorado
 ; COUNTRY: USA
 ; ZIP: 80303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/187,319
 ; FILING DATE: 27-Aug-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 09/523,656
 ; FILING DATE: 2000-03-10
 ; APPLICATION NUMBER: US 09/037,601
 ; FILING DATE: 1998-03-10
 ; APPLICATION NUMBER: WO PCT/US97/11155
 ; FILING DATE: 1997-06-26
 ; APPLICATION NUMBER: US 08/670,707
 ; FILING DATE: 1996-06-26
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Greenlee, Lorraine L.
 ; REGISTRATION NUMBER: 27,894
 ; REFERENCE/DOCKET NUMBER: 75-95K
 ; TELEPHONE: 303/499-8080
 ; TELEFAX: 303/499-8089
 ; INFORMATION FOR SEQ ID NO: 37:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2133 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 37:
 US-10-187-319-37

Query Match 81.0%; Score 6227; DB 15; Length 2133;

Best Local Similarity 57.3%; Pred. No. 0;
 Matches 1212; Conservative 106; Mismatches 119; Indels 678; Gaps 3;
 Qy 1 ATRYYILGAVELSDYMQSD-LGELPVDARPPRPVPSFPENTSVYKTLFVFTVHLF 59
 Db 20 AIRYYILGAVELSDYRQSELLRELHVDTRFAPAPCALPLGPSVLYKTVFVFTDOLF 79
 Qy 60 NIAKPRPPMGLLPTTQAEVYDVTWITLKNASHPVSLHAGVSYVWASGAYDDOTS 119
 Db 80 SVARPRPPMGLLPTTQAEVYDVTWITLKNASHPVSLHAGVSYVWASGAYDDOTS 139
 Qy 120 QREKDDKVPFGGSHYVWQVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLGALLVC 179
 Db 140 QREKDDKVPFGGSHYVWQVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLGALLVC 199
 Qy 180 REGSLAKEKTLTKHFKILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKHVTNGYVN 239
 Db 200 REGSLTREIRIQNLHEFVLLFAVFDGKSWHSETKNSLMQDRDAASARAWPKHVTNGYVN 259
 Qy 240 RSLPGLIGCHKSVYVWVIGMGTTPFVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTL 299
 Db 260 RSLPGLIGCHKSVYVWVIGMGTTPFVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTF 319
 Qy 300 LMDLGQFLLFCHISSHODGMEAYVKVDSCEEPOLRMKNEAEADYDDDLTDSMDVVR 359
 Db 320 LMDLGQFLLFCHISSHODGMEAYVKVDSCEEPOLRMKNEAEADYDDDLTDSMDVVR 378
 Qy 360 FDDNDSPSFIQIRSVAKKPKTWVHYIAAEEDMDYAPLVLPADDRSVKSYLNGPQRI 419
 Db 379 LDGDDVSPFIQIRSVAKKPKTWVHYIAAEEDMDYAPLVLPADDRSVKSYLNGPQRI 438
 Qy 420 GRKYKVRFMAYTDTFTKTRAIQHSIGILGPLLYGEVGDTLIIIFKNQASRPNIYPHG 479
 Db 439 GRKYKARFVAYTDTFTKTRAIQHSIGILGPLLYGEVGDTLIIIFKNQASRPNIYPHG 498
 Qy 480 ITDYRPLYSRRLPKGVKHLKDFPILPGEIFKYKKTVTVEGDTKSDPRCLTRYSSFNVM 539
 Db 499 ITDVSALHPGRLKLGWKLKDMPTLPGSTFKYKKTVTVEGDTKSDPRCLTRYSSFNVM 558
 Qy 540 ERDLASGLIGPLLI CYKESVDORGNQIMS DKNRNLFSVFDENRSWYLTENIQRLPMPA 599
 Db 559 EKDLASGLIGPLLI CYKESVDORGNQIMS DKNRNLFSVFDENRSWYLTENIQRLPMPD 618
 Qy 600 GVQLEDPEFQASNTMHSINGYVFDLSQLSVCLHEVAYWYIISIGAQDTDFLSVFFSGYTFK 659
 Db 619 GLQPDPEFQASNTMHSINGYVFDLSQLSVCLHEVAYWYIISIGAQDTDFLSVFFSGYTFK 678
 Qy 660 HKMYVEDTLTLFPFSGETVFMENPGLWILGCHNSDFRNRGMTALLKVSCKDKNYGY 719
 Db 679 HKMYVEDTLTLFPFSGETVFMENPGLWILGCHNSDFRNRGMTALLKVSCKDKNYGY 738
 Qy 720 EDSYEDISAVLLSKNNAIEPRSFSON----- 745
 Db 739 DNTYEDIPGILLSGKNVIEPRSFSONRPPSPASOKQFOTITSPEDDVELDPQSGERTQAL 798
 Qy 746 ----- 745
 Db 799 EELSVPDGSGLLQGNPAPHGSSSSDLQEARNEADDYLPGARERTAPSAARLRLPELH 858
 Qy 746 ----- 745
 Db 859 HSABRVLTPEPEKELKLDKMSKSSDLTKTSPITSPDLSAETERTHSLOPHHPQVNR 918
 Qy 746 ----- 745
 Db 919 SQLGAILVGNKSSHFIAGVPLGSTEDHESSLGENVSPVESDGI FEKERAHGASLTAKD 978
 Qy 746 ----- 745
 Db 979 DVLKVNLSLVKTNKARVYLKTNRKIHIDDAALTENPASATFMDKNTTASGLNHSNWI 1038
 Qy 746 ----- 745

Search completed: December 9, 2003, 17:01:37
Job time : 89 secs

Db 1039 KPLOGNPLSSRGSPPELLTSSGSGSKSVKQSSGQGRIRVAVEEELSCKGKEMMLPNSE 1098
QY 746 ----- 745
Db 1099 LTFLTNSADVQNDTHSQGKKSREEMERREKLVBQKVDLPQVYTATGKTNFLRNIFHOST 1158
QY 746 ----- 745
Db 1159 EPSVEGFDGSHAPVQDSRSLNDSABRAETHIAHFSAREAPLEAPGNRTGPPRSAY 1218
QY 746 ----- 745
Db 1219 PRVKQSLKQIRLPLEEIKPERGVVLNATSTRWSESSPILQGAKRNNLSLPFLTLEMAGG 1278
QY 746 ----- 745
Db 1279 QGKISALGKSAAGPLASGLKAVLSSAGLSEASGKAFLPKVVRHREDLLPQKTSNVSC 1338
QY 746 ----- 745
Db 1339 AHGDLQGEIFLOKTRGPVNLKNVRPGRTPSKLLGPPMPKEWESLEKSPKSTALRTKDII 1398
QY 746 -----PPVLKRHOREITRTTLOS 763
Db 1399 SLPLDRHESNHSIAAKNEGAETQREAAWTKQGGPGRCLCAPKPPVLRHRRQDISLPTFOP 1458
QY 764 DOBEIDYDITISVEMKEDPDIIYDDEBQSPRSFQKTRHYFIAAVERLWDYGMSSSPHV 823
Db 1459 EEDKMDYDITISTETKGEDFIYXGDEBQSPRSFQKTRHYFIAAVERLWDYGMSSSPRA 1518
QY 824 LRRAQSGSVQPKVKVFOEFTQSGFTOPLYRGELNEHLGLGPIYRAEVEDNIMVTFRN 883
Db 1519 LRRAQNGEVPRFKKVVREFADQSGFTQPSYRGELNKLGLGPIYRAEVEDNIMVTFKN 1578
QY 884 QASRPYSFYSSLSIYBEDQZQGAEPKKNFVKPNETKITFMKVQHMAPTKDEDFCKAWAY 943
Db 1579 QASRPYSFYSSLSIYBDDQZQGAEPKKNFVKPNETKITFMKVQHMAPTKDEDFCKAWAY 1638
QY 944 FSDVDLEKDVHSLIGLPLLVCHNTNLPALGRQVTVQEFALFPTIFDETKSWYFTENMER 1003
Db 1639 FSDVDLEKDVHSLIGLPLLVCRANTLNAHGRQVTVQEFALFPTIFDETKSWYFTENVER 1698
QY 1004 NCRAPCNIQMEDPTFKENYFHAINGYIMDTPLGLVMAQDORIRWYLLSMGSENENIHSIH 1063
Db 1699 NCRAPCHLOMEDPTLKENYFHAINGYIMDTPLGLVMAQDORIRWYLLSMGSENENIHSIH 1758
QY 1064 FSGHFTVRKKEEKYKALNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVY 1123
Db 1759 FSGHFTVRKKEEKYKAVNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVY 1818
QY 1124 SNKQOTPLGMAAGHIRFOITAGSQYQGWAPKARLHYSGSINAWSTKEPFSWIKVDLLA 1183
Db 1819 SKEQAPLGMASGRIRDFOITAGSQYQGWAPKARLHYSGSINAWSTKDPHSMIKVDLLA 1878
QY 1184 PMITHGKTGACOKFSSLYISQFIIMYSLDGKQWTVRGNSCTGLMVFFGNDVSSGIKH 1243
Db 1879 PMITHGIMTOGARQKFSLSYISQFIIMYSLDGKQWTVRGNSCTGLMVFFGNDVSSGIKH 1938
QY 1244 NIFNPPIIARYIRLHPHYTHYSIRSTLRMELMGCDLNSCMPLGMQNKASIDSDQITASSHYT 1303
Db 1939 NIFNPPIVARYIRLHPHYTHYSIRSTLRMELMGCDLNSCMPLGMQNKASIDSDQITASSHLS 1998
QY 1304 NMFATWSPSKARLHLOQGSNAWRPQVNNPKWLQVDFOKTMKVTVGTTQGVKSLTSMYV 1363
Db 1999 NIFATWSPSKARLHLOQGSNAWRPQVNNPKWLQVDFOKTMKVTVGTTQGVKSLTSMYV 2058
QY 1364 KEFLISSQDGHQWTLFPQNGKVKVQCNQDQSFPTVNNLDPPLLTLYRIRHPQSWHQI 1423
Db 2059 KEFLVSSQDGRRTWTLFLODGHQKVFQCNQDQSFPTVNNLDPPLLTLYRIRHPQSWHQI 2118
QY 1424 ALRMEVLGCEAQDLY 1438
Db 2119 ALRLEVGLCEAQDLY 2133

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 16:51:16 ; Search time 29 Seconds
(without alignments)
4768.640 Million cell updates/sec

Title: US-10-006-091-1
Perfect score: 7691
Sequence: 1 ATRRYLGAVELSDWYMSD.....WVHQIALRMEVLGCEAQLY 1438

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7227	94.0	2351	1 EZHU	coagulation factor
2	6227	81.0	2133	2 T42763	coagulation factor
3	6192	80.5	2319	2 A47004	coagulation factor
4	2390.5	31.1	2224	1 KFHUS	coagulation factor
5	2372	30.8	2211	1 KFBOS	coagulation factor
6	2369	30.8	2183	2 T42764	coagulation factor
7	1731.5	22.5	1069	1 KURU	ferroxidase (BC 1.
8	1687.5	21.9	1059	1 A35210	ferroxidase (BC 1.
9	1104	14.4	216	2 A44258	factor VIII-associ
10	663	8.6	427	2 JC4915	ags protein precu
11	657	8.5	463	1 A36479	milk fat globule m
12	650	8.5	409	2 T11743	pp47 protein - pig
13	635	8.3	401	2 S65138	glycoprotein antig
14	635	8.3	427	2 S74211	PAS-6/7 protein pr
15	543	7.1	869	2 A25945	coagulation factor
16	443	5.8	927	1 JQ9548	A5 antigen precurs
17	424.5	5.5	218	2 A47285	milk fat globule p
18	306.5	4.0	3133	2 S52093	hemocytin - silkw
19	262	3.4	845	2 JC5256	adipocyte transcri
20	221	2.9	719	2 S51739	transcription repr
21	211	2.7	1283	2 T13799	neurexin IV - frui
22	207.5	2.7	737	2 T31349	hypothetical prote
23	206.5	2.7	737	2 T15615	hypothetical prote
24	193	2.5	913	2 A48280	receptor tyrosine
25	192	2.5	876	2 A49508	protein-tyrosine k
26	185.5	2.4	855	2 S42621	protein-tyrosine k
27	185.5	2.4	910	2 A53137	tyrosine kinase re
28	173.5	2.3	819	2 I48859	tyro 10 receptor k
29	171	2.2	1381	2 T31083	paranodin - rat

30	171	2.2	1385	2 T14158	neurexin IV - mous
31	149.5	1.9	578	2 S63533	L-ascorbate oxidas
32	149	1.9	1883	2 G82875	hypothetical prote
33	148	1.9	622	2 S62580	probable multicopp
34	141.5	1.8	1196	2 S46430	botulinum neurotox
35	140.5	1.8	1196	2 J01467	toxin, nontoxic co
36	140.5	1.8	1217	2 T18209	mxg protein - Bac
37	139.5	1.8	2925	2 T00133	RNA-directed RNA p
38	139	1.8	1193	2 JC4901	nontoxic-nonhemagg
39	137.5	1.8	580	2 F84828	probable laccase (
40	137.5	1.8	903	2 F82080	preprotein translo
41	136.5	1.8	2657	2 T18497	hypothetical prote
42	135	1.8	343	2 T35030	probable copper ox
43	134.5	1.7	791	2 T16031	hypothetical prote
44	134.5	1.7	1132	2 H82887	hypothetical prote
45	133.5	1.7	567	2 T44928	L-ascorbate oxidas

ALIGNMENTS

RESULT 1

EZHU
coagulation factor VIII precursor [validated] - human
N:Alternate names: antihemophilic factor A; coagulation factor VIIIc; procoagulant co
C:Species: Homo sapiens (man)
C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text change 08-Dec-2000
C:Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S66445; I
R:Gitschier, J.; Wood, W.I.
Hum. Mol. Genet. 1, 199-200, 1992
A:Title: Sequence of the exon-containing regions of the human factor VIII gene.
A:Reference number: I54318; MUID:93265012; PMID:1303178
A:Accession: I54318
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1921,'S',1923-2351 <RES>
A:Cross-references: GB:M8648; NID:G182381; PIDN:AAA52420.1; PID:G182383
R:Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Baton, D.L.; Gitschier, J.; Keyt, B.; Seeb
Nature 312, 330-337, 1984
A:Title: Expression of active human factor VIII from recombinant DNA clones.
A:Reference number: A00525; MUID:85061548; PMID:6438526
A:Accession: A00525
A:Molecule type: mRNA
A:Residues: 1-2351 <WOO>
R:Toole, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.L.; Pittman, D.D
s', D.N.; Hewick, R.M.
Nature 312, 342-347, 1984
A:Title: Molecular cloning of a cDNA encoding human antihemophilic factor.
A:Reference number: I58059; MUID:85061550; PMID:6438528
A:Accession: I58059
A>Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-74,'V',76-1259,'E',1261-2351 <RE2>
A:Cross-references: GB:K01740; NID:G182802; PIDN:AAA52484.1; PID:G182803
R:Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.; K
.B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favaliaro, J.; Hansen, J.;
DNA 4, 333-349, 1985
A:Title: Characterization of the polypeptide composition of human factor VIII:C and t
A:Reference number: A23584; MUID:86081164; PMID:3939400
A:Accession: A23584
A:Molecule type: mRNA
A:Residues: 1-2351 <TRU>
A:Cross-references: GB:M14113; NID:G182817; PIDN:AAA52485.1; PID:G182818
R:Eaton, D.; Rodriguez, H.; Vehar, G.A.
Biochemistry 25, 505-512, 1986
A:Title: Proteolytic processing of human factor VIII. Correlation of specific cleavag
ity.
A:Reference number: A26174; MUID:86159740; PMID:3082357
A:Accession: A26174
A:Molecule type: protein
A:Residues: 20-36;392-399,'X',401-402;1668-1678;1709-1722,'D',1723-1725;1741-1755 <EA
R:Pittman, D.D.; Wang, J.H.; Kaufman, R.J.

Biochemistry 31, 3315-3325, 1992
 A>Title: Identification and functional importance of tyrosine sulfate residues within re
 A:Reference number: A42348; PMID:92207952; PMID:1554716
 A:Accession: A42348
 A:Molecule type: protein
 A:Residues: 20-36;356-371,392-408;582-594;1668-1669,'X',1671;1672-1692;1693-1708;1709-17
 A:Experimental source: recombinant material from Chinese hamster ovary cells
 A>Note: sequence extracted from NCBI backbone and corrected to correspond with the publi
 R:Pay, P.J.; Smudzin, T.M.
 J. Biol. Chem. 264, 14005-14010, 1989
 A>Title: Intersubunit fluorescence energy transfer in human factor VIII.
 A:Reference number: A43986; PMID:89340500; PMID:2503509
 A:Accession: A43986
 A:Molecule type: protein
 A:Residues: 'X',517-523;1853-1860,'X',1862-1864,'X',1866 <FAY>
 R:Leyste, A.; van Schijndel, H.B.; Niehrs, C.; Huttner, W.B.; Verbeet, M.P.; Mertens, K.;
 J. Biol. Chem. 266, 740-746, 1991
 A>Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for
 A:Reference number: A56109; PMID:91093266; PMID:1898735
 A:Contents: annotation; sulfation
 R:Gitschter, J.; Wood, W.I.; Goralka, T.M.; Wion, K.L.; Chen, E.Y.; Eaton, D.H.; Vehar,
 Nature 312, 326-330, 1984
 A>Title: Characterization of the human factor VIII gene.
 A:Reference number: A56196; PMID:85061547; PMID:6438525
 A:Contents: annotation; introns
 R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.
 Protein Sci. 4, 740-746, 1995
 A>Title: Locations of disulfide bonds and free cysteines in the heavy and light chains o
 A:Reference number: A56216; PMID:95338127; PMID:7613471
 A:Contents: annotation; disulfide bonds
 A>Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulphydryls
 R:Kjalkre, M.; Heding, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M.
 Eur. J. Biochem. 234, 773-779, 1995
 A>Title: Amino acid residues 721-729 are required for full factor VIII activity.
 A:Reference number: S63527; PMID:96163459; PMID:8575434
 A:Accession: S63527
 A:Molecule type: protein
 A:Residues: 733-752;753-759 <KJA>
 R:Lind, P.; Larsson, K.; Spira, J.; Sydow-Baekman, M.; Almstedt, A.; Gray, E.; Sandberg
 Eur. J. Biochem. 232, 19-27, 1995
 A>Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Constructio
 A:Reference number: S66445; PMID:96048024; PMID:7556150
 A:Accession: S66445
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 1668-1685 <LIN>
 C:Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure ph
 C:Genetics:
 A:Gene: GDB:F8C
 A:Cross-references: GDB:119124; OMIM:306700
 A:Map position: Xq28-Xq28
 A:Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63
 C:Function:
 A:Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro
 A:Pathway: blood coagulation
 C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
 C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-2351/Product: coagulation factor VIII #status experimental <MAT>
 F:20-740/Product: coagulation factor VIIIA heavy chain #status experimental <ACH>
 F:20-356/Domain: A1 <DA1>
 F:23-348/Domain: ferroxidase repeat homology <FO1>
 F:392-759/Domain: A2 <DA2>
 F:402-730/Domain: ferroxidase repeat homology <FO2>
 F:760-1667/Domain: B <DBO>
 F:1668-2351/Product: coagulation factor VIIIA light chain #status experimental <ACL>
 F:1709-2038/Domain: A3 <DA3>
 F:1716-2038/Domain: ferroxidase repeat homology <FO3>
 F:2039-2191/Domain: C1 <DC1>
 F:2039-2188/Domain: discoidin I amino-terminal homology <DN1>
 F:2192-2351/Domain: C2 <DC2>
 F:2192-2345/Domain: discoidin I amino-terminal homology <DN2>
 F:60,258,601,776,803,847,919,962,982,1020,1024,1074,1085,1204,1274,1278,1301,1319,1403,1

F:172-198,267-348,547-573,649-730,1851-1877,1918-1922,2040-2188/disulfide bonds: #sta
 F:355-356/Cleavage site: Arg-Met (coagulation factor Xa, protein C) #status predicted
 F:365,737,738,742,1683,1699/Binding site: sulfate (Tyr) (covalent) #status experiment
 F:391-392/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experiment
 F:414,426/Binding site: sulfate (Tyr) (covalent) #status predicted
 F:759-760/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experiment
 F:1667-1668/Cleavage site: Arg-Glu (unidentified proteinase) #status experimental
 F:1708-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experime
 F:1740-1741/Cleavage site: Arg-Ala (coagulation factor Xa) #status experimental
 F:2193-2345/disulfide bonds: #status predicted

Query Match 94.0%; Score 7227; DB 1; Length 2351;
 Best Local Similarity 61.6%; Pred. No. 0;
 Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;

QY	1	ATRRYTLGAVELSWYDMSDGLGELPVDARPPRPVPSKPPFNTSVVYKTLFVEFTVHLFN	60
DB	20	ATRRYTLGAVELSWYDMSDGLGELPVDARPPRPVPSKPPFNTSVVYKTLFVEFTVHLFN	79
QY	61	IAKPRPPWMLLGPTIOAEVDTVTITLKNMASHPVSLHVGVSVMKASEGAEYDDQTSQ	120
DB	80	IAKPRPPWMLLGPTIOAEVDTVTITLKNMASHPVSLHVGVSVMKASEGAEYDDQTSQ	139
QY	121	REKEDDKVFPGGSHYTVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR	180
DB	140	REKEDDKVFPGGSHYTVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR	199
QY	181	EGSLAKEKTTQTLHKILLFAVDFEGKSHSETKNSLMQDRDAASARAWPKMTVGVNR	240
DB	200	EGSLAKEKTTQTLHKILLFAVDFEGKSHSETKNSLMQDRDAASARAWPKMTVGVNR	259
QY	241	SUPGLIGCHRKSVYWHVIGMTTPEVHSIFLGHTFLVNRHQASLEISPIFLTAQTLL	300
DB	260	SUPGLIGCHRKSVYWHVIGMTTPEVHSIFLGHTFLVNRHQASLEISPIFLTAQTLL	319
QY	301	MDLGQFLLFCHSHQHGDGMEAYVYVDSCEPSPQLMKNNNEAEYDDDLTLDSEMDVVRP	360
DB	320	MDLGQFLLFCHSHQHGDGMEAYVYVDSCEPSPQLMKNNNEAEYDDDLTLDSEMDVVRP	379
QY	361	DDNSPSFQIRSVAKKPKTWHYITAAEEDWDYAPLVPDPRSKSYQLNNGPQIG	420
DB	380	DDNSPSFQIRSVAKKPKTWHYITAAEEDWDYAPLVPDPRSKSYQLNNGPQIG	439
QY	421	RKYKVRFWAYTDETPKTRERAOHESGILGPLYGVGDTLLIIFKNQASRPNIYPHGI	480
DB	440	RKYKVRFWAYTDETPKTRERAOHESGILGPLYGVGDTLLIIFKNQASRPNIYPHGI	499
QY	481	TDVRPLYRRLPKGVKHLKDFPILPGEIFPKYKWTVTVEDGPKSDPRCLTRYSSFWNME	540
DB	500	TDVRPLYRRLPKGVKHLKDFPILPGEIFPKYKWTVTVEDGPKSDPRCLTRYSSFWNME	559
QY	541	RLASGLIGPLLCYKESVDORGNQIMSKENVILSFVDENESWYLTENIORFLPNPAG	600
DB	560	RLASGLIGPLLCYKESVDORGNQIMSKENVILSFVDENESWYLTENIORFLPNPAG	619
QY	601	VOLEDFEQASNMIMHSINGVDFSLQLSVCLHEVAYWYILISGAQTDFLSVFSSGTYFKH	660
DB	620	VQLEDFEQASNMIMHSINGVDFSLQLSVCLHEVAYWYILISGAQTDFLSVFSSGTYFKH	679
QY	661	KMYVEDTLTLPFPGSGEIVFMENPGLWILGCHNSDFRNRGMTALLKVSQCDKNTGDYE	720
DB	680	KMYVEDTLTLPFPGSGEIVFMENPGLWILGCHNSDFRNRGMTALLKVSQCDKNTGDYE	739
QY	721	DSYEDISAVLLSKNNAIEPRSP	742
DB	740	DSYEDISAVLLSKNNAIEPRSP	799
QY	743	-----	742
DB	800	KIQNVSSDLLMLLRQSPPHGLSLDLQAKYETSDDDPSGAIQSNNSLSWTHFRPQ	859
QY	743	-----	742

860 LHSGDMVFTPEGLQLRLNEKLGTTAATTELKLDLKFVSTSTNNLIPTSDNLAAGTDN 919
743 ----- 742
920 TSSLGPPSPMPHYDSQDSTTLFGKXSPLTSGGPLSLSEBNDKSLLESLGMSQSSW 979
743 ----- 742
980 GKVSSTESGRLFKGRAHGPAALLTKONALFKVSI LLLKTNKTSNNSATNKRTHIDGPSL 1039
743 ----- 742
1040 LIENSPSYQWNI LSTDETFKVTPLIHDRMLMDKNATRLNHNKNTSSKNHMQVK 1099
743 ----- 742
1100 KEGIPDPAQNPMSFFKMLFLPESARWIQRTHGKNSLNSGGSPKQLVSLGPEKSV 1159
743 ----- 742
1160 QNFLSEKXVVVGKEFTKVLKEMVFPSSRNLF LTNLDNLHNTHNQEKKIQEEIEK 1219
743 ----- 742
1220 KETLIQENVVLPIHTVTGTNFKMKNLFLSTRQNVESYDGA YAPVLQDFRSLNDSTNR 1279
743 ----- 742
1280 TKKHTAHFKGBENLEGLNQTKQIUEKVACTTRISPNTSQNFVTVQSRKALKQPL 1339
743 ----- 742
1340 PLEETELEKRIIVDDTSTQSKNMKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI 1399
743 ----- 742
1400 PQANRSPLIATKVSFFSIRPIYLRVLFDQNSHLPAASYRKKDQSGVQESSHFLQCAK 1459
743 ----- 742
1460 NNLAILTLMTQDREVGLSTSATNSVYKVKVENTVLPKPLPTSGKVELLPKVI 1519
743 ----- 742
1520 YQDLFPTETSGSPCHLDLVEGSL LQTEGAIKWNEANRPGKVPFLRVATESAKTPSK 1579
743 ----- 742
1580 LLDPLAMDNHYGTQIPKEBWKSEKSPKTA FKKKDITLSLNACESHAAJAINBQKNP 1639
743 ----- 786
1640 EIEVTWAKQGRTERLCSQNPVLKRKHOREITRTTLQSDQEEIDYDDTISVEMKKEDFDIY 1699
787 DEDENQSPRSFOKTRHYFIAAVERLDWYGNSSSPHVLNRNQAQSVQFKKVVQFEPTD 846
1700 DEDENQSPRSFOKTRHYFIAAVERLDWYGNSSSPHVLNRNQAQSVQFKKVVQFEPTD 1759
847 GSFTQPLYRGELNHLGLLGPYIRAEVDNIMVTFRNQASRPYSYSSLSIYEEDQRCQA 906
1760 GSFTQPLYRGELNHLGLLGPYIRAEVDNIMVTFRNQASRPYSYSSLSIYEEDQRCQA 1819
907 EPRKNFVKPNETKTYFWKVOHMAPTKDFDCKAWAYFSDVDLEKDVHSLGILLVCHT 966
1820 EPRKNFVKPNETKTYFWKVOHMAPTKDFDCKAWAYFSDVDLEKDVHSLGILLVCHT 1879
967 NTLNPAHGRQVTVQEFALFTTIDETKSWYFTFENNERCRAPCNIQMEDPTFKENYRPHA 1026
1880 NTLNPAHGRQVTVQEFALFTTIDETKSWYFTFENNERCRAPCNIQMEDPTFKENYRPHA 1939
1027 INGYIMDTLPGLVMAQDQRIRWYLLSMGNSNENIHSIHFSGHVTVTRKKEEYQMALNYLP 1086
1940 INGYIMDTLPGLVMAQDQRIRWYLLSMGNSNENIHSIHFSGHVTVTRKKEEYQMALNYLP 1999

1087 GVFTVEMLPKAGIMWEVECLIGELHAGASTLFLVYSNKCQTPLGMASGHIRDFOITAS 1146
2000 GVFTVEMLPKAGIMWEVECLIGELHAGASTLFLVYSNKCQTPLGMASGHIRDFOITAS 2059
1147 GQYQWAPKARLHYSGSINAWSTKEPFSWIKVDLLAPMIHIGIKTQAGRKFSLSYISQ 1206
2060 GQYQWAPKARLHYSGSINAWSTKEPFSWIKVDLLAPMIHIGIKTQAGRKFSLSYISQ 2119
1207 FIIMYSIDGKKQWYRGSTGLMVFNGVNDSSGKINIFNPPIIARIYRLHPHYSTRS 1266
2120 FIIMYSIDGKKQWYRGSTGLMVFNGVNDSSGKINIFNPPIIARIYRLHPHYSTRS 2179
1267 TLRMELMGCDLNSCMPLGMEKSAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWR 1326
2180 TLRMELMGCDLNSCMPLGMEKSAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWR 2239
1327 PQVNNPEWLQVDFQKTMKVTVGTTQGVKSLTSMYVKEFLISSQGHQHTLFFQNGKV 1386
2240 PQVNNPEWLQVDFQKTMKVTVGTTQGVKSLTSMYVKEFLISSQGHQHTLFFQNGKV 2299
1387 KVFGQNDSTPVVNSLDPLLTXYLRHQPQSWHOIALRMEVLGCEAODLY 1438
2300 KVFGQNDSTPVVNSLDPLLTXYLRHQPQSWHOIALRMEVLGCEAODLY 2351

RESULT 2

T42763

coagulation factor VIII precursor - pig

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000

C:Accession: T42763

R:Lollar, P.

submitted to the EMBL Data Library, August 1996

A:Reference number: Z2269

A:Accession: T42763

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2133 <COL>

A:Cross-references: EMBL:U49517; NID:g1511633; PID:g1511634; PIDN:AB06705.1

C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxi-

C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; i-

F:1-19/Domain: signal sequence #status Predicted <SIG>

F:20-2133/Product: coagulation factor VIII #status predicted <MAT>

F:223-349/Domain: ferroxidase repeat homology <FOX1>

F:402-750/Domain: ferroxidase repeat homology <FOX2>

F:1498-1820/Domain: ferroxidase repeat homology <FOX3>

Query Match

Best Local Similarity 81.0%; Score 6227; DB 2; Length 2133;

Matches 1212; Conservative 106; Mismatches 119; Indels 678; Gaps 3;

Qy 1 ATRYYILGAVELSDWYMQSD-LGELPVDARPPRPVPSFPNTSVVYKTLFVETVHLF 59
Db 20 AIRYYILGAVELSDWYRQSELLRELVDTTPAPGALPLGPSLVYKTVFVEFTDOLF 79
Qy 60 NIAKPRPFWGLGPTTQAEVYDVTVTILKNMASHPSVLSHAGVSYWKASGAEYDDOTS 119
Db 80 SVARPPRPFWGLGPTTQAEVYDVTVTILKNMASHPSVLSHAGVSYWKASGAEYDDOTS 139
Qy 120 QREKDDKVPFGSGHTYVWVQLKENGPMASDPLCLITYLSYLSHVDLVKDLSGLIGALLVC 179
Db 140 QREKDDKVPFGSKQTYVWVQLKENGHTASDPLCLITYLSYLSHVDLVKDLSGLIGALLVC 199
Qy 180 REGSLAKEKTOTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKHTVNGYVN 239
Db 200 REGSLTRERTQNLHEFVLLPAVFDGKSWHSARNDSTWTRAMDPAAPAQPAHVTNGYVN 259
Qy 240 RSLPGLICCHKSKSYWVIGMTTPVHVSIFLEGHTFLVRNHRQASLEISIFLTAQTL 299
Db 260 RSLPGLICCHKSKSYWVIGMTSPVHVSIFLEGHTFLVRNHRQASLEISIFLTAQTP 319
Qy 300 LMDLGGQFLLFCHISHQHDGMEAYVKVDSCEPEEQLRMKNNEAEDYDDDDTDSEMDVVR 359

Db 320 LMDLQQLFCHISHHGGEAHRVSCAEPQLRRKADSE-EDYDDNLYDSMDVVR 378
QY 360 FDDNSPFIQIRSVAKXHPKTVWHYIAAEEEDWDYAPLAPDDRYSYKQSVLNGPQRI 419
Db 379 LDGDDVSFPIQIRSVAKXHPKTVWHYIAAEEEDWDYAPAVPSDRYSKSLYNSGPQRI 438
QY 420 GRKYKVRPMAYTDFEKTREAIQHESGILGPLLXGEVGDTLIIIFKQASRPYNIYPHG 479
Db 439 GRKYKARFVAIVDTFKRKAIPYESGILGPLLXGEVGDTLIIIFKQASRPYNIYPHG 498
QY 480 ITDVRPLSRRLPKGVKHLKDPPIIPGELFKYKWTVTVEDGPKSDPRCLTRYYSFVNM 539
Db 499 ITDVSALHGRLLKGWHLKDPPIIPGELFKYKWTVTVEDGPKSDPRCLTRYYSINL 558
QY 540 ERDLASGLIGPLLCYKESVDQRQIMSKDKRNVILFSVFDENRWYLTENQRFELPNPA 599
Db 559 EKDLASGLIGPLLCYKESVDQRQIMSKDKRNVILFSVFDENQSWYLAENQRFELPNPD 618
QY 600 GVOLEDPEFOASNIHMSINGYVFDLSQLSVCLHEVAYWYILSGACTDPLSVFPGYTPK 659
Db 619 GLQPDPEFOASNIHMSINGYVFDLSQLSVCLHEVAYWYILSVGAQTDPLSVFPGYTPK 678
QY 660 HKWYEDTLTLFPFSGETVFMENPGLWILGCHNSDFRNQMTALLKYVSSCDKNTGDIY 719
Db 679 HKWYEDTLTLFPFSGETVFMENPGLWILGCHNSDFRNQMTALLKYVSSCDRIGDIY 738
QY 720 EDSYEDISAYLSKNNAIERFSQN----- 745
Db 739 DNTYEDI PGFLLSGKNVIEPRFQAQNSRPPSASQKQFQITSPEDDDVELDPQSGERTQAL 798
QY 746 ----- 745
Db 799 BELSVPSGDGMLQNPAPHGSSSDLOEARNEADDYLPGARERNTPSAARLPELH 858
QY 746 ----- 745
Db 859 HSAERVLTEPEKELKCLDSKSSSDLLKTSFTIPSDTISAETERHSLGPPHPQVNR 918
QY 746 ----- 745
Db 919 SOLGAIVLGKNSHFTAGVPLGSTBEDHESLIGENVSPVSDGPEKERAHGPASLTKD 978
QY 746 ----- 745
Db 979 DVLKVNISLVKTNKARVYLKTNRKTHIDDAALLTENRASATFMDKNTTASGLNHVSNWI 1038
QY 746 ----- 745
Db 1039 KGPLGKNPLSSBERGPSPELLTSSGSKSVKGQSSGQGRIRVAVEEBELSGKEMMLPNSE 1098
QY 746 ----- 745
Db 1099 LTFLTNSADVQNDTHSQGKSREMERREKLVOEKVDLPQVYATGTKNFLRNIPHOST 1158
QY 746 ----- 745
Db 1159 EPSVEFGDGHAPVPQDSRLSNDASRAETHIAHFAIREAPLEAPGNRTGPGPSAV 1218
QY 746 ----- 745
Db 1219 PRVVKQLKQIRLPLEIKPERGVVLNATSTRWSESSPILQAGKRNLSLPFLTLEWAGG 1278
QY 746 ----- 745
Db 1279 QGKISALGKSAAGPLASGKLEKAVLSSAGLSGASGAEFLPKVRVHREDLLPQKTSNVSC 1338
QY 746 ----- 745
Db 1339 AHGDLQGEIFLQKTRGPNVNLKNVRPGRTPSKLLGPPMPKWNESLSPKSTALRTYKDI 1398
QY 746 -----PPVLKRQREITRTTLOS 763
|||||:|

Db 1399 SLPLDRHESNHSIAAKNEGQAETQREAAWTQGGPGRLCAPKPPVLRHRQHDISLTPQP 1458
QY 764 DOEEDIDYDITTSVMKKEDPDIYDEDRNQSPRSQKTRHYFIRAAVRELWDYGMSSSPHV 823
Db 1459 BEDKMDYDIITSTETKGEDFDIYGEDENQDPRSFKRTRHYFIRAAVEQWQWYGMSSSPPA 1518
QY 824 LRNRAQSGSVPOFKKVVQFEFTDGSFTQPLRYGBELNHLGLLGPYIRAEVEDNIMVTPRN 883
Db 1519 LRNRAQSGEVPRFKKVVPRFADGSFTQPSYRGLNKLGLLGPYIRAEVEDNIMVTPKN 1578
QY 884 QASRPYSYSLISVEEDORQASPRKRVFQNETKYTFWKVQHHMATEKBFDCAKAWAY 943
Db 1579 QASRPYSYSLISVDPDEQEQABPRHNFQVNETRYTFWKVQHHMATEDEFECDKAWAY 1638
QY 944 FSDVDLEKDVHSLIGLPLLVCHTNTLNAHQROQVTVQBFALFFTIDFDTKSWYFTENMER 1003
Db 1639 FSDVDLEKDVHSLIGLPLLCRANTLNAHQROQVTVQBFALFFTIDFDTKSWYFTENVER 1698
QY 1004 NCRAPCNTQMBDPTPKENYRFHAINGYIMDTLPLGLVMAQDQRIWYLLSMGSENENHSIH 1063
Db 1699 NCRAPCHLQMEDPTLKENYRFHAINGYIMDTLPLGLVMAQDQRIWYLLSMGSENENHSIH 1758
QY 1064 PSQHVTYRKKEEYOMALYNLYPGVFETVEMLPKAGIWRVRECLIGEHLHAGMSTFLVY 1123
Db 1759 PSQHVTYRKKEEYOMALYNLYPGVFETVEMLPKAGIWRVRECLIGEHLHAGMSTFLVY 1818
QY 1124 SNKCTPLGMAHGHTIRDFQITASGOYGQWAPKLARLHYSGSINAWSTKDPHSWIKVDLLA 1183
Db 1819 SKCQAPLGMASGRIRDFQITASGOYGQWAPKLARLHYSGSINAWSTKDPHSWIKVDLLA 1878
QY 1184 PMIHGKITQGAQKQFSSLYISQFTIMYSLDGKQWQTYRGNSTGTLMVFFGNVDSGKIH 1243
Db 1879 PMIHGKITQGAQKQFSSLYISQFTIMYSLDGKQWQTYRGNSTGTLMVFFGNVDSGKIH 1938
QY 1244 NIFNPPIIARYIRLHPTHYSIRSTLRMBELMGCDLNSCMPLGESKALSDAQITASSYFT 1303
Db 1939 NIFNPPIIARYIRLHPTHYSIRSTLRMBELMGCDLNSCMPLGESKALSDAQITASSYFT 1998
QY 1304 NMFATWSPSKARLHQGRSNARPOVNNPKELQVDFOKTKVTVGTQGVKSLTSMVY 1363
Db 1999 NIFATWSPSKARLHQGRSNARPOVNNPKELQVDFOKTKVTVGTQGVKSLTSMVY 2058
QY 1364 KEFLISSQDGHQWTLFFQNGKVKVFGNQSDSFTPVVNSLDPPLLTRYLIRHPOSWVHOI 1423
Db 2059 KEFLISSQDGHQWTLFFQNGKVKVFGNQSDSFTPVVNSLDPPLLTRYLIRHPOSWVHOI 2118
QY 1424 ALRMEVLGCEAQDLY 1438
Db 2119 ALRMEVLGCEAQDLY 2133

RESULT 3
A47004
coagulation factor VIII precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 18-Jun-1999
C:Accession: A47004
R:Elder, B.; Lajich, D.; Gitschier, J.
Genomics 16, 374-379, 1993
A:Title: Sequence of the murine factor VIII cDNA.
A:Reference number: A47004; MUID:93300511; PMID:8314577
A:Accession: A47004
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2319 <ELD>
A:Cross-references: GB:L05573; NID:g192456; PID:AAA37385.1; PID:g192457
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxid
P:1-19/Domain: signal sequence #status predicted <SIG>
P:23-349/Domain: ferroxidase repeat homology <FOI>
P:402-730/Domain: ferroxidase repeat homology <FO2>
P:1686-2006/Domain: ferroxidase repeat homology <FO3>
P:2007-2156/Domain: discoidin I amino-terminal homology <DN1>
P:2160-2313/Domain: discoidin I amino-terminal homology <DN2>

Query Match	80.5%;	Score	6192;	DB	2;	Length	2319;
Best Local Similarity	53.4%;	Pred. No.	0;				
Matches	1231;	Conservative	97;	Mismatches	103;	Indels	876;
Gaps	9;						
Qy	1	ATRYVILGAVELSWDYMOSD-LSGLPVDARPPRRVPKSPFNTSVVYKTLFVFEFTWHLF	59				
Db	20	AIRRYVILGAVELSWNVIQSDLLSVLHSDRFLPRMSTSPFNTSIMYKTVFVEYKDQLF	79				
Qy	60	NIAPRPMMGLLGPTTQAEVYDVTWITLKNMASHPVSLHAGVSVYWKASGEGABYDDOTS	119				
Db	80	NIAPRPMMGLLGPTTWTWHTVJITLKNMASHPVSLHAGVSVYWKASGEGABYDDOTS	139				
Qy	120	QREKEDKVPFGGSHYTVVWVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVC	179				
Db	140	QMEKEDKVPFGESHYTVVWVLKENGPMASDPPCLTYSYMSHVDLVKDLNSGLIGALLVC	199				
Qy	180	REGSLAKETQTLHKFTILLFAVDEGKSWHSETKNSLMQORDAASARAPMHTVNGYVN	239				
Db	200	KEGSLSKERTQMLYQFVLLFAVDEGKSWHSETNDSYQSMDSASARDWPKMHTVNGYVN	259				
Qy	240	RSPLGLIGCHRKSVYHVIQMGTTPEVHSIFLEHTEFLVRNHRQASLEISPIITFLTAQTL	299				
Db	260	RSPLGLIGCHRKSVYHVIQMGTTPEIHSIFLEHTEFFVRNHRQASLEISPIITFLTAQTL	319				
Qy	300	LMDLGQFLLFCHISSHQHGMAYVYKVDSCPEEPQLRMK-NNEEAEDYDDDLTDSMDVY	358				
Db	320	LIDLGQFLLFCHISSHKHGMAYVYKVDSCPEESQWQKNNEEMEDYDDDLV-SEMDMF	378				
Qy	359	RPDDNSPFIQIRSAVAKGPKTWVHYIAABEEDWDYAPLVAPDDRSYKSOYLNNQPOR	418				
Db	379	TLDYDSSP-FIQIRSAVAKPKTWIHYISABEEDWDYAPSVPTSDNGSYKSOYLSNGPFR	437				
Qy	419	IGRKVKVRFWAYDETFTKREAIQHESGILGPLLYGEVGDTLIIIPKNOASRPYNIYPH	478				
Db	438	IGRKVKVRFWAYDETFTKRETIQHESGLLGPLLYGEVGDTLIIIPKNOASRPYNIYPH	497				
Qy	479	GITDVRPLYSRRLPKGVKHLKDFILPCEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVN	538				
Db	498	GITDVSPLHARLPRGIKHVKDLPIHGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFIN	557				
Qy	539	MERDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLIFSVFDENRSLWLTENIQRFLPNP	598				
Db	558	PERDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLIFSVFDENRSLWLTENIQRFLPNA	617				
Qy	599	AGVQLEDPFQASIMESINGYVDSLSQVCLHEVAYWYTLTGATDPLSVFSGVTP	658				
Db	618	AKTQDPGFQASIMESINGYVDSLSQVCLHEVAYWYTLTGATDPLSVFSGVTP	677				
Qy	659	KHKMVEYEDTLTLPFSGETVPMSENPGMLTILGCHNSDFRNRMGTALLKVSSCDKNTGDY	718				
Db	678	KHKMVEYEDTLTLPFSGETVPMSENPGMLVILGCHNSDFRNRMGTALLKVSSCDKNTSDY	737				
Qy	719	YEDSYEDIYAVLLSKNNAIEPRPSQN-----	745				
Db	738	YEETIYEDIPTQLVNENNVIDPRSPFQNTNHPNTRKKFKFQDSTIPKNDMEKIEPQFERIAE	797				
Qy	746	-----	745				
Db	798	MLKVQSVSVSDMLLQGSHPTPHGLFLSDGQEAIEYRAIHDDHSPNAIDSNEGSPKVTQL	857				
Qy	746	-----	745				
Db	858	RPESHSEKIVFTPOPGLRSNKSLETTIEVKWKKLGLQVSSLPNSLMTTILSDNLKA	917				
Qy	746	-----	745				
Db	918	TFEKTDSGFPDMPVHSSSKLSTTAGFKKAYSLVGSVHPINASBENSNDILDSTLMYSQ	977				
Qy	746	-----	745				
Db	978	ESLPRDNTLSIENDRLAREKRFHGIALLTONTLTKONVSLMKTNTKTYNHTTNEKLHTE	1037				
Qy	746	-----	745				
Db	1038	SPTSIENTSTDLQDAILKVNSEIQEVTALIHDTGLGKNSYLRNLHMLNRTTSTKNKDI	1097				
Qy	746	-----	745				
Db	1098	FHRKDEBDPIQDEENTIMPFKMLFLSESSNFWKKTNGNNSLNSQEHSQKQVLYLMPFK	1157				
Qy	746	-----	745				
Db	1158	YVKNQSFLEKONKVTVEQDFTKNIGLKDMPHPNMSIFLTTLNSVHENGHRHQEKNIQE	1217				
Qy	746	-----	745				
Db	1218	EIEKEALIBEKVVLPQVHEATGSKNFLKDIILGTRQNISLYEHVHPVLQNIITSINNSTN	1277				
Qy	746	-----	745				
Db	1278	TVOIHMEHFFKRRKDKETNSBGLVNTREKVKYPSQKNITTORSKRALQOFLRSTQWLK	1337				
Qy	746	-----	745				
Db	1338	TINCSQCIKQIDHSEKMKPKFITKSSLSDSVVKSTQTNSSDSHIVKTSAPPPIDDKR	1397				
Qy	746	-----	745				
Db	1398	SPFQNKFSHVQASSYIYDFKTKSSRIQESNNFLKETKINNPSLAILPWNMFIDQKFTSP	1457				
Qy	746	-----	745				
Db	1458	GKSNSTSVYKKRENII FLKPTLPERSGKIELLPQVSIQEEIILPTSTSHSGPGLHLMK	1517				
Qy	746	-----	745				
Db	1518	EVFLQIKQPTKWKAKRCHGESIKGTESSKNTSKLLNHHAWDYHAAQIPKDMWKSKE	1577				
Qy	746	-----	751				
Db	1578	KSPETIISKQEDTILSRPHGNSHISANEKQNPQRETTWVKQGTQRTCSQIIPVVKR	1637				
Qy	752	HQREITRTTLQSDQEEIDYDITISVEMKEDFDIYDENQSPRSFOKTRHYPIAAVER	811				
Db	1638	HQREL--SAPQSOEATDYDAITIE--TIEDFDIYSEDIKQSPRSFOKTRHYPIAAVER	1694				
Qy	812	LWYDGMSSPHVLRNRAQSGSVQFKVVFQEFDTGSGFTQPLVRGELNEHLGLGPYIRA	871				
Db	1695	LWYDGMSTS-HVLNRVQSDNVQFKVVFQEFDTGSGFTQPLVRGELNEHLGLGPYIRA	1753				
Qy	872	EVEDNIMVTFRNQAARPSYSSLI SYEEDOROQAEPRKPVKPNETKTYFWKVQHMAP	931				
Db	1754	EVEDNIMVTFRNQAARPSYSSLI SYEEDOROQAEPRKPVKPNETKTYFWKVQHMAP	1812				
Qy	932	TKDEPDCANAYSDVDLEKDVHSGILGPLLACHTNTLNPAGHGOVTVQEFALFTFDE	991				
Db	1813	TEDEFDCANAYSDVDLEKDVHSGILGPLLACHTNTLNPAGHGOVTVQEFALFTFDE	1872				
Qy	992	TKSWYFTENNERCRAPCNIQMEDPTPKENYRFAHNGYIMDTLPLGVMAQDQIRWYLL	1051				
Db	1873	TKSWYFTENNERCRAPCNIQMEDPTPKENYRFAHNGYIMDTLPLGVMAQDQIRWYLL	1932				
Qy	1052	SMGNSNENIHSFSGHVFTVRKKEEYKALYNLYPGVFETVEMLPKAGIWRVECLIGE	1111				
Db	1933	SMGNNENIQSIHFSGHVFTVRKKEEYKALYNLYPGVFETVEMLPKAGIWRVECLIGE	1992				
Qy	1112	LHAGMSTFLVYNNKCOPTPLGMASGHRDPOITASGOYQWAPKLARLHYSGSINASTK	1171				
Db	1993	LQAGMSTFLVYNNKCOPTPLGMASGHRDPOITASGOYQWAPKLARLHYSGSINASTK	2052				
Qy	1172	EPFSWIKVDLLAPMI IHGKITQAGAROKFSSLYTSQFIIMYSLDGKKWOTYRGSTGTLMV	1231				
Db	2053	EPFSWIKVDLLAPMI IHGKITQAGAROKFSSLYTSQFIIMYSLDGKKWOTYRGSTGTLMV	2112				
Qy	1232	FFGNVDSSGKHNIFNPPIIARIYIRLHPTHYSIRSTLRMBELMGCDLNSCMPLGWESKAI	1291				

Db 2113 PFGVNDSSGKHNSENPRIIYIRLPHTHSSIRSTRMLWGLDNLSCSIPLGWESKVI 2172
 Qy 1292 SDAQITASSYFTNMPATWSPKARHLQGRSHNAPQVNNPKENIQLVDFOKTMKVTGTT 1351
 Db 2173 SDTQITASSYFTNMPATWSPQARHLQGRSHNAPQVNNPKENIQLVDFOKTMKVTGTT 2232
 Qy 1352 QGVKSLTSMYKVEFLISSQDGHOWTLFFQNGKVKVFGQNDSTFPVNSLDPPLLTXY 1411
 Db 2233 QGVKSLTSMYKVEFLISSQDGHOWTLFFQNGKVKVFGQNDSTFPVNSLDPPLLTXY 2292
 Qy 1412 LRIHPQSWHQIALRMEVLGCEAQLY 1438
 Db 2293 LRIHPQSWHQIALRMEVLGCEAQLY 2319

RESULT 4

KFHUS

coagulation factor V precursor [validated] - human
 N:Alternate names: coagulation labile factor; proaccelerin
 C:Species: Homo sapiens (man)
 C>Date: 19-May-1989 #sequence revision 02-Jun-1995 #text_change 08-Dec-2000
 C:Accession: A56172; A42344; A28028; A27498; A25897
 R:Cripe, L.D.; Moore, K.D.; Kane, W.H.
 Biochemistry 31, 3777-3785, 1992
 A:Title: Structure of the gene for human coagulation factor V.
 A:Reference number: A42344; MUID:92232668; PMID:1567832
 A:Accession: A56172
 A:Molecule type: DNA
 A:Residues: 48-58;79-89;120-130;191-201;239-249;313-323;368-378;428-437;461-471;533-542; 2070;2111-2120;2172-2181 <CR2>
 R:Jenny, R.J.; Pittman, D.D.; Toole, J.J.; Kriz, R.W.; Aldape, R.A.; Hewick, R.M.; Kaufm Proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987
 A:Title: Complete cDNA and derived amino acid sequence of human factor V.
 A:Reference number: A28028; MUID:87260886; PMID:3110773
 A:Accession: A28028
 A:Molecule type: mRNA
 A:Residues: 1-857,'R',859-864,'R',866-924,'E',926-1763,'I',1765-2212,'T',2214-2224 <JEN>
 A:Cross-references: GB:M16967
 A:Note: Parts of this sequence, including the amino end of the mature protein, were determined by protein sequencing
 R:Kane, W.H.; Ichinose, A.; Hagen, F.S.; Davie, E.W.
 Biochemistry 26, 6508-6514, 1987
 A:Title: Cloning of cDNAs coding for the heavy chain region and connecting region of human factor V.
 A:Reference number: A27498; MUID:88107560; PMID:2827731
 A:Accession: A27498
 A:Molecule type: mRNA
 A:Residues: 1-1284,'I',1286-1600 <KAN>
 A:Cross-references: GB:M17785
 A:Note: Parts of this sequence were determined by protein sequencing
 R:Kane, W.H.; Davie, E.W.
 Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986
 A:Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor homologous to factor VIII.
 A:Reference number: A25897; MUID:86313665; PMID:3092220
 A:Accession: A25897
 A:Molecule type: mRNA
 A:Residues: 1188-1215,1315-2224 <KA2>
 A:Cross-references: GB:M14335
 A:Note: Parts of this sequence were determined by protein sequencing
 R:Keller, F.G.; Ortal, T.L.; Quinn-Allen, M.A.; Kane, W.H.
 Biochemistry 34, 4118-4124, 1995
 A:Title: Thrombin-catalyzed activation of recombinant human factor V.
 A:Reference number: A56139; MUID:95210278; PMID:7696276
 A:Contents: annotation; thrombin cleavage sites
 C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
 C:Genetics:
 A:Gene: GDB:F5
 A:Cross-references: GDB:1119896; OMIM:227400
 A:Map position: 1q23-1q23
 A:Introns: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 65

C:Function:
 A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa p1
 A:Pathway: blood coagulation
 C:Superfamily: coagulation factor V; discooidin I amino-terminal homology; ferroxidase
 C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasme
 F:1-28/Domain: signal sequence #status predicted <SIG>
 F:19-224/Product: coagulation factor V #status predicted <MAT>
 F:29-737/Product: coagulation factor V heavy chain #status experimental <VAH>
 F:29-345/Domain: A1 <DA1>
 F:33-329/Domain: ferroxidase repeat homology <FO1>
 F:346-691/Domain: A2 <DA2>
 F:351-684/Domain: ferroxidase repeat homology <FO2>
 F:692-1573/Domain: B <DOB>
 F:1183-1461/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
 F:1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
 F:1574-1905/Domain: A3 <DA3>
 F:1581-1905/Domain: ferroxidase repeat homology <FO3>
 F:1667-1765/Region: phospholipid binding #status predicted
 F:1906-2064/Domain: C1 <DC1>
 F:1906-2061/Domain: discooidin I amino-terminal homology <DN1>
 F:2065-2224/Domain: C2 <DC2>
 F:2065-2221/Domain: discooidin I amino-terminal homology <DN2>
 F:51-55,239,297,460,468,554,741,752,760,776,782,821,938,977,1074,1083,1103,1106,1479;
 F:167-193,248-329,500-526,603-684,1725-1751,1907-2061,2066-2221/Disulfide Bonds: #sta
 F:334-335/Cleavage site: Arg-Asn (protein C) #status predicted
 F:363,693,1546/Binding site: sulfate (Tyr) (covalent) #status predicted
 F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
 F:382,1338/Binding site: carboxylate (Asn) (covalent) #status absent
 F:534-535/Cleavage site: Arg-Gly (protein C) #status predicted
 F:737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:1046-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experimental
 F:1573-1574/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match 31.1%; Score 2390.5; DB 1; Length 2224;
 Best Local Similarity 26.1%; Pred. No. 6.4e-151;
 Matches 593; Conservative 277; Mismatches 483; Indels 923; Gaps 35;

Qy 3 RRYLGAVELGWDYVQSDGLGELPVDPRPVPKSPFNTSVV-YKTLFVFTVHLFNI 61
 Db 32 RQFYVAAQGISWYRPE-----PTNSLNLSTVSPKKIVREYEPY-PKK 75
 Qy 62 AKRPPNMGILGPTTQARVYDTVTITLKNMASHVSLHAGVYVYKASGEGADYDQTSOR 121
 Db 76 EKQPTISGLLPTLYAEVGDIIKVFKNKADKPLSIHPQGIYRSKLSGASYLDHTFPA 135
 Qy 122 EKEDDKVPFGSHYVQVLEKNGPMASDPCLTYSYLSHVDLVKOLNSGLIGALIVCEE 181
 Db 136 EKMDDAVAPGREYTYVENSISDSGPTDDPPCLTHIYSHENLIEDFNSGLIGLILCKK 195
 Qy 182 GSLAKEKTO-TLHK-FILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYV 239
 Db 196 GLTLGGTQKTFDQIVLLFAVDESKSWSSQSS-----LMTVNGYV 239
 Qy 240 RSLPGLIGCHRSVYVHVGMGTTPVHSHIFLEGHTFLVRNHRQASLEISPTFTTAQTL 299
 Db 240 GTMPDITVCAHDHISWLLGSSGPELFSIHFGQVLEQNHKVSATILVSATSTANMT 299
 Qy 300 LMDLQFLLFCHISSHQHDGMEAVYKVDSCPEEPQLRMKNNEAEYDLDLTDSEMDVR 359
 Db 300 VGFEKWIISLTPKHLQAGWQAYIDKNCPKKTKNLKKTRE----- 342
 Qy 360 FDDNDSPIQIRSVAKKPKTWVHYIAAEEDWDYAPLVLPADDDRSYKSOYINNGPORI 419
 Db 343 -----QRRMKWEVFIAAEEVWDYAPVIANMDKYSQSHLDNFSNQI 387
 Qy 420 GRKYKVRMYATDFTFKTREAI-----QHEGILGPLLYGEGVDTLLIIFKNQASRPYNI 476
 Db 388 GRHYKVMYQYDEDSF-TKHTVNPMMKEDGILGPIIRAQVRDTLKIYFKNMASRPYSI 446
 Qy 477 PHGIT-----DVRFLYSRLPKGVKHLKDFPILGGEIFKYKWTYVEDGPTKSPRCLT 530
 Db 447 PHGVTFSPYEDVNGSFT-----SGRNTMIRAVQGETTYTKWNILEDFDEPTENDACCLT 502

QY 531 RYSSFNVERDASGLGPIII CYKESVDQGNQIMSDKRNVLPSVDENRSLTEN 590
Db 503 RPYSDVDIMDIAAGLLGLLICKSRSDRRGIQRAADI BOQAVFAVDENKSWYLEDN 562
QY 591 IQREFLPAGVQLEDPEPQASNMHSINGYVFDLSL-QLSVCLHEVAYVYVILSIGAQTDPL 649
Db 563 INKFCENDEVKRDDPKYESNIMSTINGYPESITTLGFCFDDIVQHFCSVGTQNEIL 622
QY 650 SVFSGYTPKHMVYEDTLTPFSGETVFMENPGLWILGCHNS----- 695
Db 623 TIHPTGHSFIYKGRHEDTLTPFMGESVTVMNDVGTWMLTSMNSSPRSKKLRLKPRDV 682
QY 696 -----DPENR----- 700
Db 683 KCIPDDSDSYEIFEPPSTVMATRKMDRLPEDEDESADYDQNRLLAALGIRSPRNS 742
QY 701 -----GMTAL----- 705
Db 743 SINQEEEFNLTALENGTEFVSNTDII VGSNYSSPSNISKFTVNNLAEPOKAPSHQ 802
QY 706 -----LKVSSCDKNTGYDEDSYED----- 725
Db 803 ATTAGSPURHLIGKNSVLNSTAHS SPYSEDPIDPLQPDVGTGIRLLSLGAGBFSQEH 862
QY 726 ----- 725
Db 863 AXHKGPVKVERDQAACHRESWMKLLAHKVGRLHSQDTSPSGMRPWEDLFSQDTSPSRMR 922
QY 726 -----ISAYLLSKNA----- 736
Db 923 PKDPPSLLLLKQNSKILVGRWHLASEKGSVEIIQDTEDETRAVNNWLLSPQNASRAW 982
QY 737 ----- 736
Db 983 GESTPLANKPGQSGHPKPRVRHKSQVRQDGGKSLKKSQFLIKTRKKKKKHETHAP 1042
QY 737 IEPSPF-----SQ 744
Db 1043 LSPRTFFPLRSEAYNTFSERLLKHSVLVHKSNETSLPTDLNQTLPMSDFGHIASLPDHQ 1102
QY 745 N----- 745
Db 1103 NSSNDTQASCPGLGYQVTPPEEHYQTPPIQDPDQMHSTSDPSHRSSSPELSEMLEYDRS 1162
QY 746 -----PVLKXH----- 752
Db 1163 HKSPPTDISQSPSEHEVQTVISPDLSQVTLSPELSQTNLSPDLSHTTSLSPELIQRL 1222
QY 753 ----- 752
Db 1223 SPALGQMPIPDLSHTTSLSPDLSHTTSLDLSQTNLSPELSQTNLSPALGQMLSPDLSH 1282
QY 753 ----- 752
Db 1283 TTLSLDFSQTNLSPELSHMTLSPELSTQTNLSPALGQMPIPDLSHTTSLDLSQTNLSPE 1342
QY 753 -----QREITRTTLOSQDOEBI----- 768
Db 1343 LSQTNLSPALGQMLSPDPSSHHTTSLDLSQTNLSPELSTQTNLSPDLSMPFLADLSQIPL 1402
QY 769 ----- 768
Db 1403 TPDLDQMTLSPDLGETDLSPNFGQMSLSPDLSQVTLSPDISDTTLLPDLSQISPPDLQ 1462
QY 769 -----DYDDT 773
Db 1463 IFYPSESQSLLOFENESFPYDPLGQMPSPSPPLNDTFLSKFNPLIVVCLSKDGRDY 1522
QY 774 ISVEMKKK-----DFDIYDE-----DENQSP-----RSFOKKTRHY 804
Db 1523 IEIIPKEVQSEDDYABIDIVPDYDDPKYKTVRTNINSSRPDNTAAWYLSNNGNRNRY 1582
QY 805 FIAAVERLWDYGMSSPHVLRNR--AQSGSVFQ----FKKVVVFQEFDTGDSFTQPLYRGLN 859

Db 1583 YIAAEISWDY-----SEFVQRETIEDSDDDIPEDTTAKYVFRKYLSTFTKRDPRGRYE 1638
QY 860 EHLGLLGPYIRAEVEDNIMVTRNQASRPYGFYSLSISYE-----EDORQCAEPKRN 911
Db 1639 EHLGLLGPYIRAEVDDVIVQVFRKNLSPYSLHAHGLSYEKSSEKTKYVDEDDSPFWFKEDN 1698
QY 912 FVKNETKTYFWKQVQHMAPTKDBFDCKAMAYFSDVLEKDVHSLGILGIPLLVCHTNTLNP 971
Db 1699 AVQPNSSYTYVWHATERSGPSPGSCACRAMAYSAVNPEDKIDHSGILGIPLLICQKGI LHK 1758
QY 972 AHGQVTVQEFALPFTLFDETKSNYFTENMERNCRAPCNIQWEDPTFKENTRFHAINGYI 1031
Db 1759 DSNPLDMREFVLLFMFTFDEKKSYWEKKSSWR-----LTSEMKKSHFEHAINMI 1812
QY 1032 MDTLPLGLVMAQDQIRWYLLSMGSENINHSIHFSGHVFTVRKKEGYSKALNLYLPGVPE 1091
Db 1813 Y-SLPLGLKMYEQEWRVRLHLLNIGSQSDIHVVHFGQTLLENGNKQHQGLGWPLLP 1871
QY 1092 VEMLPKAGIWRVRECLIGEHLAGMSTLFLVYSNKCQTPPLQMASGHIRDFOITASGOYQ 1151
Db 1872 LEMKASKPGWMLLATEVGENQORAGWQTPFLIMDRDCRMPMGLSTGIISDSQIKASEFLGY 1931
QY 1152 WAPKLARLHYSGSINAMSTKE---PFS---WIKVDLAPMIIHGIKTCGAKROKFSLSYIS 1205
Db 1932 WEPLARLNNGSGYNASWEKLAAPFASKPWIOVDMQKEVIITGIOTQGAHYLKSCTYT 1991
QY 1206 QFIIMYLDGKKWQTYRGNSGTLMVFFGNVDSGIGHNIENPIIARIYIRLHPHYISIR 1265
Db 1992 EFYVAYSSNOINWQIFKGNSTRVWYFNGSDATIKENQDPDPPIVARIIRISPRAYNR 2051
QY 1266 STLRLMELGCDLNSCMLPGMESKAISDAQITASSYFTNMFEAT-WSPEKARLHLQGRSNA 1324
Db 2052 PTLRLQLQCEVNGCSTPLGMENKIKENQITASSFKKSWMGDYWEPEARLNAQGRVNA 2111
QY 1325 WRPOVNPKEWLOVDFOKTMKVTVGTTCGVKSLTSMYKFLISSSQDGHQWTLFFQNG 1384
Db 2112 WQAKANNKOWLEIDLKIKKIKITAITITQGCCKSLSEMYKSYTYTHYSQGVKPYRLKS 2171
QY 1385 KV--KVFQGNQDSFTPVVNSLDPLLTLYLRIHPQSWHQAIALRMEVLGCEAQDLY 1438
Db 2172 SWVDKIFEGNTYTKGHVKNFNFPIISRFIKVIPTKWNQSLAKLLELFGC---DIY 2224

RESULT 5

KFBO5

coagulation factor V precursor - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999

C:Accession: A42580; A36497

R:Guinto, E.R.; Eamon, C.T.; Mann, K.G.; MacGillivray, R.T.

J. Biol. Chem. 267, 2971-2978, 1992

A:Title: The complete cDNA sequence of bovine coagulation factor V.

A:Reference number: A42580; MUID:92147638; PMID:11737753

A:Accession: A42580

A:Molecule type: mRNA

A:Residues: 1-2211 <GUI>

A:Cross-references: GB:W81440; NID:g163037; PIDN:AAA30512.1; PID:g163038

A:Note: sequence extracted from NCBI backbone (NCBI:80774, NCBI:80776)

R:Kalafatis, M.; Jenny, R.J.; Mann, K.G.

J. Biol. Chem. 265, 21580-21589, 1990

A:Title: Identification and characterization of a phospholipid-binding site of bovine

A:Reference number: A36497; MUID:91072354; PMID:2254316

A:Accession: A36497

A:Molecule type: protein

A:Residues: 1566-1570, 'X', 1572-1581, 'X', 1583-1584, 1673-1676, 'X', 1678-1679, 'X', 1681, 'X',

R:Xue, J.; Kalafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.

Biochemistry 33, 13109-13116, 1994

A:Title: Determination of the disulfide bridges in factor Va heavy chain.

A:Reference number: A55979; MUID:95034740; PMID:7947716

A:Contents: annotation

A:Note: 566-Cys and 617-Cys were shown to have free sulphydryls

C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.

Db 1581 SWDYS-----KVSDDVDVYPTVYKVKVFRKILSDSTFKLPDQSEYEHGIL 1631
Qy 866 GPYIRAEVEDIMVTRNOARPPSYFSLISYE-----EDORQGAERPKNFVKE 917
Db 1632 GPVIRAEVDDVQVFRKSLRPPYSLAHAGLSYKSSRGKTYEDDSDPEWFKEDNAIQNK 1691
Qy 918 TKTYFWKHQWMAPTKDFEDCKAMAYFSDVDLEKDVHGLGIPLLVCHVTNINPAHQGV 977
Db 1692 TTYTYVHAATRSRGPENPGSACRWAYYSVAVNPEKDIHSLGIPLLICRGLTLDKTNVP 1751
Qy 978 TVQSFALPFTLFDKTSVYFTEENMERNCRAPCNIQWEDPTKENVREHAINGYMDTLPG 1037
Db 1752 DMRFEVLLFWFDEKSYIYDKKPKTRSWRRASS-----EYKNSHEFAHNGMIYN-LPG 1804
Qy 1038 LVMAQDQRIRWYLSMGSNENIHSIHFGSHVTRVRKEEYKALYNLYPGVFETVEMLPS 1097
Db 1805 LRMYEQEVRVRLHLLNGLGSRDHHVHFHGTLLNGTQHQQLGVWPLLPSSFKTLEMCAS 1864
Qy 1098 KAGIWRVCECLIGEHLAGHAGTILFLVYSNKQTPPLGMSGHIRDPOITASQVQWAPKLA 1157
Db 1865 KPGWMLLDTVEGEIQRAGMOTPFLLVDRCKMPMLSTGLIADSIQASBFMGWEPKLA 1924
Qy 1158 RLHYSGSINAW-----STK-BPFSWKVYDILLAPMIIHGKIQGARQKFSLYISQFIWY 1211
Db 1925 RLNCGSSNAMAIEKLSFENPEPWIQDMQKEVLLTGITQGAHYLKPYITTEFCVAY 1984
Qy 1212 SLDGKKWQTYRGNSGTGLMWFGNVDSGIGKHNIENFPPIIARIYRLHPTHYSIRSLAME 1271
Db 1985 SLDRKNWIRFGNSRNVWYFGNSDASTIKENQIDPPVARYIRISPTGSYNKPALRE 2044
Qy 1272 LMGCNLNCSPLMGESKAISDAQITASSYFTNNPAT-WSPSKARLHIOGNSNWRPOVN 1330
Db 2045 LQCEVNGCSTPLMGESKIKENQITASSFKSMWNGYWEPPFARLINAQGRVNAQAKAN 2104
Qy 1331 NPKEWLQVDFQTKWKTGVTQGVKSLTSMVKEFLSSQDGHOWTLFQNGKV--KV 1398
Db 2105 NNQWLQIDLLKIKITAIYVQGCSSLSSEMYVKSYYTHYSDQGDWPKPYREKSSMVDKI 2164
Qy 1389 FQGNQDSFTPVVNSLDPPLLYRLIHPQSWHQAIALMEVLGCEAODLY 1438
Db 2165 FEGNNVYRGVKNFNPFIISRFRIIPKTNQSLALELFGC---DMY 2211
RESULT 6
T42764
coagulation factor V - mouse
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
C:Accession: T42764
R:Yang, T.L.; Cui, J.; Rehmtulla, A.; Yang, A.; Moussalli, M.; Kaufman, R.J.; Ginsburg, R.; Blood 91, 4593-4599, 1998
A:Title: The structure and function of murine factor V and its inactivation by protein C
A:Reference number: 222270; MUID:98282202; PMID:9616155
A:Accession: T42764
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2183 <YAN>
A:Cross-references: EMBL:U52925; NID:G3219690; PID:G3219691; PIDN:AAC99553.1
C:Function:
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;
F:350-662/Domain: ferroxidase repeat homology <FOX1>
F:1541-1864/Domain: ferroxidase repeat homology <FOX2>
Query Match 30.8%; Score 2369; DB 2; Length 2183;
Best Local Similarity 25.8%; Pred. No. 1.7e-149;
Matches 577; Conservative 287; Mismatches 486; Indels 888; Gaps 31;
Qy 3 RRYVLGAVELSWDYNQSDGLGELPVADRPFRPVKSPFPNTSVVYKTLFVETVHLFNIA 62
Db 32 RQFYVAAQGLIWNHYE-----PTDPSLNSIPSPFKIYVREYQY-FKKE 75

Qy 63 KPRPPMGLIGPTTOAEVYDVVITLKNMASHPVSLHAGVSYWKAASGAEYDDQTSORE 122
Db 76 KPRSSNSGLLPTLYAEVGVKIVHFRNKADKPLSIHPQGIKYSKFSGASYADHTPPAE 135
Qy 123 KEDDKVPKGSHTYVWVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCREG 182
Db 136 RKDAPVAPGEYTYEVIWVSDESGPTDDPPCLTHIYYSYENLTQDFNSGLPLICKKG 195
Qy 183 SLAKEKTQTL--HKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVR 240
Db 196 TLTEGDTQKMPDKXHVLLFAVFDSEKS-----RSQSPSLMYTINGFVYK 239
Qy 241 SLPLGICHRKSVYVHVGMTTPEVHSIFLEHTFLVRNHRQASLEISPTFTTAOTLL 300
Db 240 TMPDITVCAHDHVSMLHIGMSSGPELFSIHFGVQVLEQNHQKSVTVLVSATSTANMTM 299
Qy 301 MDLQQLFLPCHISHQHDGMEAYVKVDSCEPEPOLRMKNNEEAEDYDDDLTDSMDVVRP 360
Db 300 SPGRWIVSSLIPIHYQAGMQAYIDIKNCPKCTR----- 333
Qy 361 DDDNSPFIQIRSVAKKHPKPTWVHYIAAEEDWDYAPLVLPADDDRSYKSOVLNNGPORIG 420
Db 334 ----SPK--TLTRQRRYMKWEYFIAAEVINYAPVIPANMDKIYRSQHLDFNSQIG 387
Qy 421 RYKVKVRFMAYTDETP--KTREAIQHSIGILGPLLYGEVGTLLIIFKNQASRPYNIYPH 478
Db 388 KHYKVIYRQVBEETFTKRTDNPSIKQSGILGPVIRAQVRDLKAIYFNKASRPYSIYPH 447
Qy 479 GITDVRPLSRRLPKGV-----KHLKDPILPGEIYKYYKWTVTVEDGPTKSPDRCLTR 531
Db 448 GVT-----FSPYEDGINSSTSGSHTTIRPVQPGFTFYKWNILEFDEPTENDAQCILTR 501
Qy 532 YSSFVNMERDLASGLIGPLLCYKESVDQGRNOIMSKRNVILFSYVDENRSMYLTNI 591
Db 502 PYSDVDVTRDIASGLIGLLIICKRSLDQGRVADIEQQAFAVAFVDENKSYIEDNI 561
Qy 592 QRLFPNPAGVQLEDPFQASNIMHSINGYVDFSQ--LSVCLHEVAYVYILSIGAOTDPLS 650
Db 562 NKCPENDEVKRDDPKFVESNIMSTINGYVPESITLGFCDPTVQMHFCSVGHDDILT 621
Qy 651 VFFSGYFKHKWYEDTLTLPFSGETVFMNSMENPGLWILGCHNSDFNRGMATALLKVSS 710
Db 622 IHFTGHSFIYGRREDTLTLFPMRGESVTVMQNVGTWMLTTMNSPKRRLRLRFRDVK 681
Qy 711 CDKNTGY-VEDSYE-----DISAYLLS----- 732
Db 682 CNR---DYDNEDSYEIEPPAPTSMTTRIHDSLENEFGIDNEDDDYQYLLASSLGRSF 738
Qy 733 KNAIEPR----- 740
Db 739 KNSLNFEENFNLTALALENSSEFFISPTDVRVDSNSSRSLSKIINNKLDFORTLPGS 798
Qy 741 ----- 740
Db 799 GATVAGTLRLNLIGLDENFVINSSTHRSSTSYHENDMENPQSNITWVLLPLGPKGSNR 858
Qy 741 ----- 740
Db 859 EQDKPKTIKTRGPRHMKHRFSWMAKAPAGKTRHSNPNKNSYSGMKSEBEDIPELIPLKQKI 918
Qy 741 ----- 740
Db 919 TSKFLNRRWRVASEKGSVEIIAANGEDTDVDKLTNSPQONQITVPRGESTSHNTTRKPS 978
Qy 741 ----- 740
Db 979 DLPTFSGVGHKSPHVRQEEENSGFKQKLFTRTKKKKKKKLALHSLSPRGPDLRGHN 1038
Qy 741 ----- 740
Db 1039 HSPFPDRRLNLHLLHKSNETALSPDLNQTSPSMSTRSLPDLVQYKNDTEQMSSSLD 1098

QY 741 -----SFSQNPVVKR----- 751
Db 1099 LYQSVPAEHSPTTFAQPDQTHSTTDPYSRSPPELSQGLDYLDHDFYDDIGLTSFF 1158
QY 752 ----- 751
Db 1159 PQSQKSPSSDDQAIPSSDLSLFTISPDLQTIIPDQLLSPEDNQKTSPPDLQ 1218
QY 752 ----- 751
Db 1219 VPLSPDDNQKTSPPDLQGVLSFPDDNQKTSPPDLQGVLSLDDNQKTSPPDLQVPLSPD 1278
QY 752 -----EIRTTTQSQE----- 767
Db 1279 DNQMITSPDLQGVPLSSDNQKTSPPDLQGVPLFPEDNQKTYFLDSQVPLSSDNQKTSST 1338
QY 768 ----- 767
Db 1339 DLLTLSPDPQTVLSPDLQPLPDSNQVTVSPDLSLLTLSPDFNEIILAPDLQVTLIS 1398
QY 768 ----- 767
Db 1399 PDLIQTNPALNHGKASSADPPQASYPDPGQASSLPFLNRLPHPLPHIPPPSPPTL 1458
QY 768 -----IDYDDT-----ISVMKKEDPDYEDENQSPRSFQ 798
Db 1459 NNTSLSRKENPLVVGLSRVGDGVEIIVPSEPERIDEDYAEDDFVYNDPYRTDTRDV 1518
QY 799 KTRH-----YFTAAVERLWDYGMSSPHVLNRRAQSGVPO-----FK 837
Db 1519 NSSRPDPTTAAWYLRGHGKFKFYIAEEITWYAEFAQSEM--DHEDTGTPTKDTYTK 1576
QY 838 KVFQEPDGTSTQPLRYGELNEHLGLGYIYRAEVEDNIMTFRQASRPYSFVSSLIS 897
Db 1577 KVFYKYLSTSTPRPRAEYBEHLGILGFVIRAEEVDVIVQRFKMLASRPYSLHAHGLS 1636
QY 898 YE-----EPQRCQAEPRKQVFNENKTYFKVQHHMAPTQDFDCKAWAYSDDVL 949
Db 1637 YKSSSEKTYEDESPEMFQEDDAVQPNSSYTVVWHTKSGENPSCACRANAYYSANV 1696
QY 950 EKDVHSGLIGPLLVCHTNTINPAHQGVQTVVQSFALFTTFIDETKSYWFTENMERNCRAPC 1009
Db 1697 ERDIHSLIGLPLLCRKGTLERNLPMDBREFVLLFWFDEKKSYYEKS--KGSR--- 1751
QY 1010 NQEMEDPTKENYRFHAIINGYIMDTPLGLVMAQDQIRWYLLSMGSENIHHSCHVF 1069
Db 1752 --RIBSPEEKNAHKFYAINGMIYN-LPGLRMVEQEWVRLHLLMGGSRDIHVHFGQTL 1808
QY 1070 TVRKKEEYKVALYNLYPGVFETVEMLPKSGAGIWRVECLICEHLHAGMSTLFLVYNNKQCT 1129
Db 1809 LDNRVYKQHLGVWPLLPGSFKTLEMKASKPGWLLDTEVGENQVAGMQTPFLIIDKECKM 1868
QY 1130 PLGMAHGIRDPQITASGYGOWAPKLARLHYSGSINAWSTKE-----PFS-WIKVDLLA 1183
Db 1869 PMGLSTGVLSDSQIKASELYLTWEPLRLNAGSYNWSIEKTDLPDIKWIQVDMQK 1928
QY 1184 PMIHGKTCGARQKESLYISQFIITMYSIDGKQWYTGNSGTGLMVFFGNVDSGKIKH 1243
Db 1929 EYVVTGICQGAHYLKSCTFTEFQVAYSDDQTNWQIFRGKSGKSYMYFTGNSDGTIKE 1988
QY 1244 NTFNPPIIARYIRLHPHYTHYSIRSLRMELMGCDLNSCPLMGESKAISDAQITASSYFT 1303
Db 1989 NRLDPPIVARYIRIHFTKSYNRTFLLELQCEVNGCSTPLGLEDGRIQDKQITASSFPK 2048
QY 1304 NMFAT--WSPSKARLHLQGRSNARWPPQVNNPKLEQLVDFOKTMKVTVGTTQGVKSLTSMY 1362
Db 2049 SWMGDYWEPSLARLNAQGRVNAQKANNKQMLQVLDLLKIKVTAIVTQGCSSLSSEMY 2108
QY 1363 VKFEFLISSQDGHQWTLFPQNGKV--KVFQGNQDSFTPVVNSLDPLFLRYLRIHQSVW 1420
Db 2109 VYSKSIQYSDQGVAMKPYRKQKSMVDKIFEGNSNTKGMKNFFNPPIISRFTIIRIPKTN 2168
QY 1421 HQIALRMEVLGCEAQDLY 1438

Db 2169 QSIARLELFGC---DIY 2183
RESULT 7
KUUH
ferroxidase (EC 1.16.3.1) precursor [validated] - human
N:Alternate names: ceruloplasmin
N:Contains: ferroxidase long form (CP-1); ferroxidase short form (CP-2)
C:Species: Homo sapiens (man)
C:Date: 31-Aug-1980 #sequence revision 12-May-1995 #text_change 08-Dec-2000
C:Accession: A25443; A24165; A35450; A00524; I59067
R:Koshinsky, M.L.; Funk, W.D.; van Oost, B.A.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 5086-5090, 1986
A:Title: Complete cDNA sequence of human ceruloplasmin.
A:Reference number: A25443; MUID:86259737; PMID:2873574
A:Accession: A25443
A:Molecule type: mRNA
A:Residues: 1-1060,1065-1069 <KOS>
A:Cross-references: GB:M13699; NID:g180255; PIDN:AAA51976.1; PID:g180256
A:Note: this is the short or CP-2 alternatively spliced form
R:Mercer, J.F.B.; Grimes, A.
FEBS Lett. 203, 185-190, 1986
A:Title: Isolation of a human ceruloplasmin cDNA clone that includes the N-terminal 1
A:Reference number: A24165; MUID:86275241; PMID:3755405
A:Accession: A24165
A:Molecule type: mRNA
A:Residues: 1-40;549-599;784-829;919-952 <MER>
R:Yang, F.; Friedrichs, W.B.; Cupples, R.L.; Bonifacio, M.J.; Sanford, J.A.; Horton, J.
J. Biol. Chem. 265, 10780-10785, 1990
A:Title: Human ceruloplasmin. Tissue-specific expression of transcripts produced by a
A:Reference number: A35450; MUID:90285218; PMID:2355023
A:Accession: A35450
A:Molecule type: DNA
A:Residues: 1007-1064 <YAN>
A:Cross-references: GB:J05506
A:Note: this is the long or CP-1 alternatively spliced form
R:Takahashi, N.; Ortel, T.L.; Putnam, F.W.
Proc. Natl. Acad. Sci. U.S.A. 81, 390-394, 1984
A:Title: Single-chain structure of human ceruloplasmin: the complete amino acid sequen
A:Reference number: A00524; MUID:84119493; PMID:6582496
A:Accession: A00524
A:Molecule type: protein
A:Residues: 20-1060,1065-1069 <TAK>
A:Note: 79-Gly and 449-Gly were also found
R:Yang, F.; Naylor, S.B.; Lum, J.B.; Cutshaw, S.; McCombs, J.L.; Naberhaus, K.H.; McG
Proc. Natl. Acad. Sci. U.S.A. 83, 3257-3261, 1986
A:Title: Characterization, mapping, and expression of the human ceruloplasmin gene.
A:Reference number: I59067; MUID:86205876; PMID:3486416
A:Accession: I59067
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 218-1069 <RES>
A:Cross-references: GB:M13536; NID:g180248; PIDN:AAA51975.1; PID:g180249
C:Comment: Ferroxidase is a blue, plasma alpha2-glycoprotein binding 6-7 copper ions
C:Comment: In Wilson's disease the plasma levels of ferroxidase are diminished or und
C:Comment: The three fragment chains are produced spontaneously during purification a
C:Genetics:
A:Gene: GDB:CP
A:Cross-references: GDB:I19069; OMIM:I17700
A:Map position: 3q23-q25
A:Introns: 1006/3; 1061/1
A:Note: the list of introns is incomplete
C:Function:
A:Description: catalyzes the oxidation of free iron(II) to iron(III) coupled with the
A:Note: iron(III), but not iron(II), is the form bound and transported by transferrin
A:Note: other possible functions are amine oxidase activity, copper transport and hom
C:Superfamily: ferroxidase; ferroxidase repeat homology
C:Keywords: acute phase; alternative splicing; copper; duplication; glycoprotein; oxi
F1/19/Domain: signal sequence #status predicted <SIG>
F20-1069/Product: ferroxidase, long form #status predicted <MALT>
F20-1060/Product: ferroxidase, short form #status experimental <MATS>
F20-499/Product: ferroxidase 67K chain #status experimental <K67>

Query Match	22.5%	Score 1731.5	DB 1	Length 1069
Best local Similarity	32.6%	Pred. No. 2.4e-107		
Matches	385	Conservative 195	Mismatches 409	Indels 193
Gap#	22			
QY	3	RRYLGAVELSWDVMQSDLGS--LPVDARPPRPVKSPFNTSVVYKKTLLFVETVHLF	59	
DB	22	KHYTIGIETWDY-ASDHGKKLISVDTEHSNLYLQNGPDRIQRLYKAUQLQTTDET	80	
QY	60	NIAXPRPPMGLLGTPTQAEYDVTVTILKONMASHPSLHAGVSYWKASGASBYDDQTS	119	
DB	81	RTTIEKFWLGFGLPIIKAEATGDKVYVHLLKNLASRPYTFHSHGITYYKHEGALYPDNWT	140	
QY	120	QREKEDKVPFGGSHTYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALIVC	179	
DB	141	DFRADDKVPGEQYTYTMYLLATEQSPGEGDNCVTRIYHSHIDAPKDIASGLIGPLTIC	200	
QY	180	REGSLAKEKQTL-HKFTLLFAVDEGKSHSE-----TKNSLMQRPDASARAWPMH	232	
DB	201	KQOSLDKBKKHIDREFVWFVSVDENFSWYLEDNIKTYCSPEKVDKONEDFOESNRMY	260	
QY	233	TVANGYVNRSLPGLTGCHRKSYVHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPT	292	
DB	261	SVNGYTGSLPGLSNCAEDRVKWTFLGNGNEVDVHAFFHQALTKNRYRDTINLFPAT	320	
QY	293	FLTAQTLMLDGLQFLFLCHSSHQHDGMEAYKVYDSCPEEPQLRMKNNEEAEDDDDLTD	352	
DB	321	LFDAYVQAQNGEWMLSQNLNHLKAGLQAFFQVQEC-----NKSSSKD-----	364	
QY	353	SEMDVVPDDNSPSFTQIRSVAKKHPTWVHYIAAHEEDWDYAPL-----VLAPD	403	
DB	365	-----NIRGKHVRH-----YYIAAEEIWNYPASGIDIFTKENLTAPG	402	
QY	404	DRSVKSOVLNNGPQIRGKYKKVFMAYTDTF--KTREAIQESGLGLPLLYGEGVDT	460	
DB	403	SDS--AVPFGCTTRIGSYKLYRYETDASFTNPKERGEPEEHLGLGLGVIAEWGDT	460	
QY	461	LLIIFKNQASRYNINYPHGI-----TOVRPLY--SRRLPGVKHLKDFPILPGEIF	509	
DB	461	IRVTFHNKGAYPLSIEPIGVRFNPKNNEGTYTSPNYPQSRVPPSASH-----VAPTET	515	
QY	510	KYKWTIVVEDGPTSKDPRCLTRYSSFVNMRDLASGLIGPLLYCYKESVDQRGNQMSD	569	
DB	516	TYEWTVPKEVGFTWADPVCIAKTYSAVDPTKIFTGLIGEMKICKGSLHANGRKQDND	575	
QY	570	KRNVLSPVDENRSWYLTENIQRLPNPAGVQLEDPEFOASNTMHSINGVYVFSLQ-LS	628	
DB	576	KBFYLPFTVDENESLLEDNIRMTTAPQVDKEDDFQSNKKHNSGNGMYGNQPGLT	635	
QY	629	VCLHEVAYWYILSLCAQTDFLSVPFSGVYTFKKVMVYEDTLTLFPFSGETVFMWSMENGLW	688	
DB	636	MCKGDSVWVWYLFSAAGNEADVHGIYFSGNWTYLWRGERDTANLFPQTSILTHMWPDTGTF	695	
QY	689	ILGCHNGDFRNRGMTALKVSSCDKNTGDYVEDYSAYLLSKNNAIERSPSFQNPVPV	748	
DB	696	NVECLTDDHYTGGMKQKTVYVQACRRQS-----EDS-----	725	
QY	749	LKRHQREITRTLQSDQBEIYDDTISVEMKKEPFDIYDENQSPRSFQKTBHYFTAA	808	
DB	726	-----TFYLGERTYYIAA-----	738	

F;1-21/Domain: signal sequence #status predicted <SIG>
F;28-60/Domain: EGF homology <EG1>
F;68-107/Domain: EGF homology <EG2>
F;110-267/Domain: discoidein I amino-terminal homology <DN1>
F;271-427/Domain: discoidein I amino-terminal homology <DN2>

Query Match 8.6%; Score 663; DB 2; Length 427;
Best Local Similarity 40.3%; Pred. No. 1.6e-36;
Matches 142; Conservative 61; Mismatches 123; Indels 26; Gaps 6;

QY 1105 ECLIGELHAGMSTLFLVYSNK-----CQPLGWSGHIRDFOITASGQY-- 1149
DB 78 KCLVTDTRQ--DFTYIICQCPVYSGIHCELCGCKTGLEGAIDQSISAVYMG 135
QY 1150 ----GQWAPKLARLHYSGINAW--STKEPFSWKVLDLLAPMIHGIKTQARQKFSLSY 1203
DB 136 FMGLQRWGPPELRLRYTGIVNWTASSYDSKFWLQVDFLRKRVSGVMTQASRAGAEY 195
QY 1204 ISQFIIMYSLDGKKWQTVRGNSGTGLMVFFGVNDSSGKHNIFNPPIIARYIRLHPHYS 1263
DB 196 LKTEKVAISLDGRREFEFTQDESCTGDKFEFGNQDNNSLKINWFNFTLEAQYIRLVPVSC 255
QY 1264 IRSLRMELMGCDLNSCMPLGMSKAISSDAQITASSYFT--NMFA-TWSPSKARLHLOG 1320
DB 256 RGCTLRFLGCEHLGCGSEPLGLKNNITIPDSQITASSSYKTWNLRFAFGWYPHLGRDNQ 315
QY 1321 RSNARWQVNNPKWQLQVDFQKTMKVTVTTQGVKSLTSMYVKEFLISSQDGHQWTLF 1380
DB 316 KINAWTAQNSAKWELQVLDLTKQKVTGIITQAGDFGHQIVASYKVAHSDGQVMTVY 375
QY 1381 FQNGKVKFQNGQSDFTVPVNSLDPPLTRYLRHQPQSWHQAIALRMEVLGC 1432
DB 376 BEQTSKVFQGNLDNNSHKKNIFEKPFMARYVRVLPFLSWHNRITLRLLELGC 427

RESULT 11

A36479
milk fat globule membrane protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A36479
R;Stubbs, J.D.; Lekutis, C.; Singer, K.L.; Bui, A.; Yuruki, D.; Srinivasan, U.; Parry, C.
Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990
A;Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the ex
A;Reference number: A36479; MUID:91046008; PMID:2122462
A;Accession: A36479
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-463 <STU>
C;Cross-references: GB:M38337; NID:g199142; PIDN:AAA39534.1; PID:g199143
C;Superfamily: milk fat globule protein; discoidein I amino-terminal homology; EGF homol
C;Keywords: membrane protein
F;68-107/Domain: EGF homology <EG1>
F;147-303/Domain: discoidein I amino-terminal homology <DN1>
F;307-463/Domain: discoidein I amino-terminal homology <DN2>

Query Match 8.5%; Score 657; DB 1; Length 463;
Best Local Similarity 42.2%; Pred. No. 4.7e-36;
Matches 135; Conservative 63; Mismatches 110; Indels 12; Gaps 5;

QY 1124 SNKQTPLGWSGHIRDFOITASGQY-----GQWAPKLARLHYSGINAW--STKEPFS 1175
DB 145 ASRCSTQLGMEGAIDQSISAVYMGFLQRWGPELRLRYTGIVNWAHSNYSK 204
QY 1176 WIKVDLLAPMIHGIKTQARQKFSLSYISQFIIMYSLDGKKWQTVRGNSGTGLMVFFGN 1235
DB 205 WIQVNLKRMVSGVMTQASRAGAEYLKTFKVAISLDGRKFEFIQDESQGD-KBFLGN 263
QY 1236 VDSGKHNIFNPPIIARYIRLHPHYSIRSLRMELMGCDLNSCMPLGMSKAISSDAQ 1295
DB 264 LDNNSLKNVNFNFTLEAQYIRLVPVSCRGCTLRFLGCEHLGCEPLGLKNNITIPDSQ 323

QY 1296 ITASSYFT--NMFA-TWSPSKARLHLOGRSNAMRPQVNNPKWQLQVDFQKTMKVTVTTQ 1352
DB 324 MSASSSYKTWNLRFAFGWYPHLGRDNQKINAWTAQNSAKWELQVLDLGTQRTQVTGIIT 383
QY 1353 GVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKVKVFGQHQDSFTVPVNSLDPPLTRYL 1412
DB 384 GARDFGHIOYVESYKVAHSDGQVMTVYBQSSKVFQGNLDNNSHKKNIFEKPFMARYV 443
QY 1413 RIHPQSWHQAIALRMEVLGC 1432
DB 444 RVLPSWHNRITLRLLELGC 463

RESULT 12

T11743
pp47 protein - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 08-Sep-2002
C;Accession: T11743
R;Ensslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Toei
Biol. Reprod. 58, 1057-1064, 1998
A;Title: Molecular cloning and characterization of P47, a novel boar sperm-associated
A;Reference number: 217325; MUID:98206817; PMID:9546740
A;Accession: T11743
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-409 <ENS>
A;Cross-references: EMBL:Y11683; NID:g2652927; PIDN:CAA72379.1; PID:g2652928
A;Experimental source: testis
C;Function:
A;Description: may be involved in membrane remodeling and/or function as a zona pelluc
C;Superfamily: milk fat globule protein; discoidein I amino-terminal homology; EGF hom
F;6-40/Domain: EGF homology <EGF>

Query Match 8.5%; Score 650; DB 2; Length 409;

Best Local Similarity 39.3%; Pred. No. 1.1e-35;
Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

QY 1105 ECLIGELHAGMSTLFLVYSNK-----CQPLGWSGHIRDFOITASGQY-- 1149
DB 58 ECEVIDDAHRG--DVFTYEICKPHGYTGHCCEIICNAPLOMETGAIDQISASSWHLG 115
QY 1150 ----GQWAPKLARLHYSGINAW--STKEPFSWKVLDLLAPMIHGIKTQARQKFSLSY 1203
DB 116 FMGLQRWAPELARLHRAGINAWTASVNDPMPQVNLRRMRTVGVVTVQASRAGAEY 175
QY 1204 ISQFIIMYSLDGKKWQTVRGNSGTGLMVFFGVNDSSGKHNIFNPPIIARYIRLHPHYS 1263
DB 176 MKTFKVAISYDGRKRFQIQAGBSGDKIFMGNLDNSGLKYNLFEPVLEVQVRLVPIICH 235
QY 1364 IRSLRMELMGCDLNSCMPLGMSKAISSDAQITASSYFTN--MFATWSPSKARLHLOG 1320
DB 236 RGCTLRFLGCELSGCGEPLGLKNNITIPKQITASSFYRTWGLSAPSWFFPYARLNDQ 295
QY 1321 RSNARWQVNNPKWQLQVDFQKTMKVTVTTQGVKSLTSMYVKEFLISSQDGHQWTLF 1380
DB 296 KFNAWTAQNSASEWLQIDLGSQRRVTGIITQAGDFGHQIVYAAVKVAISDDGVSTWEY 355
QY 1381 FQNGKVK--KVPQNGQSDFTVPVNSLDPPLTRYLRHQPQSWHQAIALRMEVLGC 1432
DB 356 RDQGALEGGKIFPGNLDNNSHKKNIFEKPFRLTFVRILPVAWHNRITLRLLELGC 409

RESULT 13

S65138
glycoprotein antigen MGp57/53, mammary gland - bovine (fragment)
N;Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E8
C;Species: Bos primigenius taurus (cattle)
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
C;Accession: S65138; G48394
R;Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A;Title: Molecular cloning of glycoprotein antigens MGp57/53 recognized by monoclonal

A;Reference number: S65138; MUID:96125736; PMID:8541316

A;Accession: S65138
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-401 <AOK>
R;Mather, I.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A;Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig II-like sequences.
A;Reference number: A48394; MUID:93250576; PMID:8485470
A;Accession: G48394
A;Status: preliminary
A;Molecule type: protein
A;Residues: 207-220 <MAT>
R;Experimental source: milk
A;Note: sequence extracted from NCBI backbone (NCBIP:131457)
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology <DN1>
F;1-32/Domain: EGF homology (fragment) <EG1>
F;82-239/Domain: discoidin I amino-terminal homology <DN2>
F;743-401/Domain: discoidin I amino-terminal homology <DN2>

Query Match 8.3%; Score 635; DB 2; Length 401;

Best Local Similarity 37.9%; Pred. No. 1.1e-34;

Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 1105 ECLGEHLHAGNSTLFLVYSNK-----CQPLGNASGHIRDFQITASGQY-- 1149

DB 50 ECQVTDSDHRG--DVFIQYICKPLGYVGHICETCTSPFGMQTGAIADSOISASMHG 107

QY 1150 ----GQWAPKLARLHVYSGINAWST----KEPFSWIKYDILAPMIHGIKQGARQKFS 1201

DB 108 FNGLQWAPELARLHQTGLVNAWTSQNDKNP--WIQVNLKMKWVTGVVTCASRAGSA 165

QY 1202 LYISQFIIMYSLDGKKWQYRGNSTGLTMVFFGNVDSGSIKHNFPPPIIARYLRHPTH 1261

DB 166 EYLKTFKVAYSTDGRQFQIQAAGSGDKIFIGNVNSGLKINLPDTPLEQYVRLVPII 225

QY 1262 YSIRSLRMELMGCDLNSCMPGLMESKAISDAQITASSYFTN---MFATWSPSKARLHL 1318

DB 226 CHRGTCLFELLGCELCNGTEPLGLKNDTIPNKQITASSYKWTGLSAFSPFPYYARLDN 285

QY 1319 QGRSNAPQVNNPKWELQVDQKTMKVTGVTQGVKSLTSMYKEFLISSQDGHWT 1378

DB 286 QOKFNWTAQTNASAEWLQIDLSQKRVGTGIIQGARDFGHQYVAARVAYGDDGVVT 345

QY 1379 LFFQNG--KVKVFGNQDSFTPVVNSLDPPLLRILYLRHPSQVHQAIALRMEVLGC 1432

DB 346 EYKDPGASESKIFPGMNDNSHKKNIFETFPQARFVRIOQVAVHNRITLRLVLLGC 401

RESULT 14

S74211
PAS-6/7 protein precursor - bovine
N;Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E8
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Dec-1997 #sequence revision 12-Dec-1997 #text_change 04-Nov-2002
C;Accession: S74211; S78114; S24181; S65138; G48394
R;Hvarregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.
Eur. J. Biochem. 240, 628-636, 1996
A;Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat globule
A;Reference number: S74211; MUID:97008954; PMID:8856064
A;Accession: S74211
A;Molecule type: mRNA
A;Residues: 1-427 <HVA>
A;Cross-references: EMBL:X91895; NID:g1632778; PIDN:CAA62997.1; PID:g1632779
A;Accession: S78114
A;Molecule type: protein
A;Residues: 19-85;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;425-427
R;Kim, D.H.; Kanno, C.; Mizokami, Y.
Biochim. Biophys. Acta 1122, 203-211, 1992
A;Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, from

A;Reference number: S23926; MUID:92353107; PMID:1643094

A;Accession: S24181
A;Molecule type: protein
A;Residues: 383-394 <KIM>
R;Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Mateuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A;Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal
A;Reference number: S65138; MUID:96125736; PMID:8541316
A;Accession: S65138
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 27-427 <AOK>
R;Mather, I.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A;Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig II-like sequences.

A;Reference number: A48394; MUID:93250576; PMID:8485470

A;Accession: G48394
A;Status: preliminary
A;Molecule type: protein
A;Residues: 233-246 <MAT>
R;Experimental source: milk
A;Note: sequence extracted from NCBI backbone (NCBIP:131457)
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
C;Keywords: blocked amino end; disulfide bond; glycoprotein; milk
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-427/Product: PAS-6/7 protein #status experimental <MAT>
F;24-58/Domain: EGF homology <EG1>
F;66-105/Domain: EGF homology <EG2>
F;108-285/Domain: discoidin I amino-terminal homology <DN1>
F;269-427/Domain: discoidin I amino-terminal homology <DN2>
F;24-35,29-47,49-58,66-77,71-94,96-105/Disulfide bonds: #status predicted
F;27/Binding site: carbohydrate (Ser) (covalent) #status experimental
F;34/Binding site: carbohydrate (Thr) (covalent) #status experimental
F;59,227/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;109-265,252-256,270-427/Disulfide bonds: #status experimental

Query Match 8.3%; Score 635; DB 2; Length 427;

Best Local Similarity 37.9%; Pred. No. 1.2e-34;

Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 1105 ECLGEHLHAGNSTLFLVYSNK-----CQPLGNASGHIRDFQITASGQY-- 1149

DB 76 ECQVTDSDHRG--DVFIQYICKPLGYVGHICETCTSPFGMQTGAIADSOISASMHG 133

QY 1150 ----GQWAPKLARLHVYSGINAWST----KEPFSWIKYDILAPMIHGIKQGARQKFS 1201

DB 134 FNGLQWAPELARLHQTGLVNAWTSQNDKNP--WIQVNLKMKWVTGVVTCASRAGSA 191

QY 1202 LYISQFIIMYSLDGKKWQYRGNSTGLTMVFFGNVDSGSIKHNFPPPIIARYLRHPTH 1261

DB 192 EYLKTFKVAYSTDGRQFQIQAAGSGDKIFIGNVNSGLKINLPDTPLEQYVRLVPII 251

QY 1262 YSIRSLRMELMGCDLNSCMPGLMESKAISDAQITASSYFTN---MFATWSPSKARLHL 1318

DB 252 CHRGTCLFELLGCELCNGTEPLGLKNDTIPNKQITASSYKWTGLSAFSPFPYYARLDN 311

QY 1319 QGRSNAPQVNNPKWELQVDQKTMKVTGVTQGVKSLTSMYKEFLISSQDGHWT 1378

DB 312 QOKFNWTAQTNASAEWLQIDLSQKRVGTGIIQGARDFGHQYVAARVAYGDDGVVT 371

QY 1379 LFFQNG--KVKVFGNQDSFTPVVNSLDPPLLRILYLRHPSQVHQAIALRMEVLGC 1432

DB 372 EYKDPGASESKIFPGMNDNSHKKNIFETFPQARFVRIOQVAVHNRITLRLVLLGC 427

RESULT 15

A25945
coagulation factor VIII - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 16-Aug-1988 #sequence revision 16-Aug-1988 #text_change 21-Jul-2000
C;Accession: A25945
R;Toole, J.J.; Pittman, D.D.; Orr, E.C.; Murtha, P.; Wasley, L.C.; Kaufman, R.J.

Proc. Natl. Acad. Sci. U.S.A. 83, 5939-5942, 1986
A;Title: A large region (approx 195 kDa) of human factor VIII is dispensable for in vitro
A;Reference number: A25945; MUID:86287369; PMID:3016730
A;Accession: A25945
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-869 <TOO>
C;Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase

Query Match 7.1%; Score 543; DB 2; Length 869;
Best Local Similarity 75.2%; Pred. No. 5.4e-28;
Matches 100; Conservative 1%; Mismatches 19; Indels 0; Gaps 0;
QY 746 PPVLRHQRHRTTTLQSDQEEIDYDITISVEMKKEDFDIYDEDENQSPRSFQKTRHYF 805
DB 737 PPVLRHQRDISLPTRPQEDKMDYDIFSTETKGEDFDIYGEDENQDPRSFQKTRHYF 796
QY 806 IAAVERLWDYGMSSSPHVLNRAQSGSVPOFKVVFQEFDTQSGFTQPLYRGELNEHLGLL 865
DB 797 IAAVEQLWDYGMSSSPALNRAQNGEVPRFKVFRERADGSFTNPSYRGELNKHGLL 856
QY 866 GPYIRAEVEDNIM 878
DB 857 GPYIRAEVEDNIM 869

Search completed: December 9, 2003, 16:55:42
Job time : 57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 16:49:51 ; Search time 18 seconds
(without alignments)
3756.911 Million cell updates/sec

Title: US-10-006-091-1

Perfect score: 7691

Sequence: 1 ATRRYLGAVELSWDMQSD.....VVHQIALRMEVLGCEAQDLY 1438

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	7227	94.0	2351	1	FAB_HUMAN	P00451	homo sapien
2	6227	81.0	2133	1	FAB_PIG	P12263	sus scrofa
3	6192	80.5	2319	1	FAB_MOUSE	Q06194	mus musculus
4	2389.5	31.1	2224	1	FAS_HUMAN	P12259	homo sapien
5	2372	30.8	2211	1	FAS_BOVIN	Q28107	bos taurus
6	2357.5	30.7	2258	1	FAS_PIG	Q9GLP1	sus scrofa
7	1727.5	22.5	1065	1	CERU_HUMAN	P00450	homo sapien
8	1684.5	21.9	1059	1	CERU_RAT	P13635	rattus norv
9	1635.5	21.3	1062	1	CERU_MOUSE	Q61147	mus musculus
10	663	8.6	427	1	MFGM_RAT	P70490	rattus norv
11	657	8.5	463	1	MFGM_MOUSE	P21956	mus musculus
12	650	8.5	409	1	MFGM_PIG	P79385	sus scrofa
13	635	8.3	427	1	MFGM_BOVIN	Q95114	bos taurus
14	588	7.6	387	1	MFGM_HUMAN	Q08431	homo sapien
15	469.5	6.1	931	1	NRP2_HUMAN	O60462	homo sapien
16	464.5	6.0	925	1	NRP2_RAT	Q03276	rattus norv
17	462.5	6.0	931	1	NRP2_MOUSE	O35375	mus musculus
18	458.5	6.0	914	1	NRP1_CHICK	P79795	gallus gall
19	451.5	5.9	922	1	NRP1_RAT	Q9QWJ9	rattus norv
20	446.5	5.8	923	1	NRP1_MOUSE	P97333	mus musculus
21	443	5.8	928	1	NRP1_XENLA	P28824	xenopus lae
22	429.5	5.6	923	1	NRP1_HUMAN	O14786	homo sapien
23	306.5	4.0	3133	1	HMC2_BOMMO	P98092	bombyx mori
24	266	3.5	764	1	CPX2_MOUSE	Q94215	mus musculus
25	261	3.4	756	1	CPX2_HUMAN	Q9H436	homo sapien
26	260.5	3.4	280	1	XLR1_FUGRU	Q9W6R5	figu rubrip
27	247	3.2	224	1	XLR1_MOUSE	Q9Z114	mus musculus
28	243	3.2	224	1	XLR1_HUMAN	O15537	homo sapien
29	223	2.9	722	1	CPXM_MOUSE	Q92100	mus musculus
30	217	2.8	734	1	CPXM_HUMAN	Q965m3	homo sapien
31	217	2.8	1331	1	CTA2_HUMAN	Q9unc6	homo sapien
32	211	2.7	1284	1	NRX4_DROME	Q94887	drosophila
33	208	2.7	1310	1	CTA4_MOUSE	Q99p47	mus musculus

ALIGNMENTS

```

RESULT 1
FAB_HUMAN
ID - FAB_HUMAN          STANDARD;          PRT; 2351 AA.
AC P00451;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Coagulation factor VIII precursor (Procoagulant component)
DE (Antihemophilic factor) (AHF).
GN F8 OR F8C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=86081164; PubMed=3935400;
RA Truett M.A., Blacher R., Burke R.L., Caput D., Chu C., Dina D.,
RA Hartog K., Kuo C.H., Masiaz F.R., Merryweather J.P., Najarian R.,
RA Pachi K., Potter S.J., Puma J., Quiroga M., Rall L.B., Randolph A.,
RA Urdea M.S., Valenzuela P., Dahl H.-H.M., Favalaro J., Hansen J.,
RA Nordfang O., Ezban M.;
RT "Characterization of the polypeptide composition of human factor
RT VIII:C and the nucleotide sequence and expression of the human kidney
RT cDNA.";
RL DNA 4:333-349(1985).
RN [2]
RP MEDLINE=85061548; PubMed=6438526;
RA Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschier J.,
RA Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wion K.L.,
RA Delwart E., Tuddenham E.G.D., Vehar G.A., Lawn R.M.;
RT "Expression of active human factor VIII from recombinant DNA clones.";
RL Nature 312:330-337(1984).
RN [3]
RP MEDLINE=85061550; PubMed=6438528;
RA Toole J.J., Knopf J.L., Wozney J.M., Soltzman L.A., Buecker J.L.,
RA Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C.,
RA Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Fass D.N.,
RA Hewick R.M.;
RT "Molecular cloning of a cDNA encoding human antihemophilic factor.";
RL Nature 312:342-347(1984).
RN [4]
RP MEDLINE=93265012; PubMed=1303178;
RA Gitschier J., Wood W.I.;
RT "Sequence of the exon-containing regions of the human factor VIII
RT gene.";
RN Hum. Mol. Genet. 1:199-200(1992).
RN [5]
RP SEQUENCE OF 2064-2070 FROM N.A.
RA de Water N.S., Williams R., Browett P.J.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SULFATION OF TYR-1699.

```

```

34 206 2.7 1308 1 CTA4_HUMAN
35 198.5 2.6 854 1 DDR2_MOUSE
36 194 2.5 1288 1 CTA3_HUMAN
37 192 2.5 913 1 DDR1_HUMAN
38 185.5 2.4 855 1 DDR2_HUMAN
39 185.5 2.4 910 1 DDR1_RAT
40 185.5 2.4 911 1 DDR1_MOUSE
41 183 2.4 1384 1 CTA1_HUMAN
42 171 2.2 1381 1 CTA1_RAT
43 171 2.2 1385 1 CTA1_MOUSE
44 149.5 1.9 578 1 ASO_TORAC
45 148 1.9 622 1 YAK3_SCHPO

```

```

Q9c0a0 homo sapien
Q62371 mus musculus
Q9bz76 homo sapien
Q08345 h epithelia
Q16832 homo sapien
Q63474 rattus norv
Q93146 mus musculus
P78357 homo sapien
P97846 rattus norv
O54991 mus musculus
Q40588 nicotiana t
Q09920 schizosacch

```

RX MEDLINE=91093266; PubMed=1898735;
 RA Leyte A., van Schijndel H.B., Niehrs C., Huttner W.B., Verbeet M.P.,
 RA Mertens K., van Mourik J.A.;
 RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is
 RT essential for the interaction of factor VIII with von Willebrand
 RT factor.";
 RL J. Biol. Chem. 266:740-746(1991).
 RN [7]
 RN SULFATION.
 RP MEDLINE=92207952; PubMed=1554716;
 RA Pittman D.D., Wang J.H., Kaufman R.J.;
 RT "Identification and functional importance of tyrosine sulfate
 RT residues within recombinant factor VIII.";
 RL Biochemistry 31:3315-3325(1992).
 RN [8]
 RN STRUCTURE BY NMR OF 2322-2343.
 RP MEDLINE=95200924; PubMed=7893714;
 RA Gilbert G.E., Baleja J.D.;
 RT "Membrane-binding peptide from the C2 domain of factor VIII forms an
 RT amphipathic structure as determined by NMR spectroscopy.";
 RL Biochemistry 34:3022-3031(1995).
 RN [9]
 RN REVIEW ON MOLECULAR BASIS OF HEMA.
 RP MEDLINE=91221499; PubMed=1902642;
 RA Gitschier J.;
 RT "The molecular basis of hemophilia A.";
 RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
 RN [10]
 RN REVIEW ON MOLECULAR BASIS OF HEMA.
 RP MEDLINE=89088506; PubMed=2491949;
 RA White G.C. II, Shoemaker C.B.;
 RT "Factor VIII gene and hemophilia A.";
 RL Blood 73:1-12(1989).
 RN [11]
 RN REVIEW ON MOLECULAR BASIS OF HEMA.
 RP MEDLINE=95245332; PubMed=7728145;
 RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
 RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
 RL Hum. Mutat. 5:1-22(1995).
 RN [12]
 RN VARIANT HEMA GLN-2326.
 RP MEDLINE=86225434; PubMed=3012775;
 RA Gitschier J., Wood W.I., Shuman M.A., Lawn R.M.;
 RT "Identification of a missense mutation in the factor VIII gene of a
 RT mild hemophiliac.";
 RL Science 232:1415-1416(1986).
 RN [13]
 RN VARIANT HEMA PRO-2135.
 RP MEDLINE=88096539; PubMed=3122181;
 RA Levinson B., Janco R.L., Phillips J.A. III, Gitschier J.;
 RT "A novel missense mutation in the factor VIII gene identified by
 RT analysis of amplified hemophilia DNA sequences.";
 RL Nucleic Acids Res. 15:9797-9805(1987).
 RN [14]
 RN VARIANT HEMA GLN-2228.
 RP MEDLINE=88191889; PubMed=2833855;
 RA Yousoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
 RA Kazazian H.H.;
 RT "Nonsense and missense mutations in hemophilia A: estimate of the
 RT relative mutation rate at CG dinucleotides.";
 RL Am. J. Hum. Genet. 42:718-725(1988).
 RN [15]
 RN VARIANT HEMA GLY-291.
 RP MEDLINE=88220354; PubMed=2835904;
 RA Yousoufian H., Wong C., Aronis S., Platokoukis H., Kazazian H.H. Jr.,
 RA Antonarakis S.E.;
 RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
 RT in exon 7 of the factor VIII gene.";
 RL Am. J. Hum. Genet. 42:867-871(1988).
 RN [16]
 RN VARIANT HEMA CYS-1708.
 RP MEDLINE=89274393; PubMed=2499363;
 RA O'Brien D.P., Tuddenham E.G.;

RT "Purification and characterization of factor VIII 1,689-Cys: a
 RT nonfunctional cofactor occurring in a patient with severe hemophilia
 RT A.";
 RL Blood 73:2117-2122(1989).
 RN [17]
 RN VARIANT HEMA CYS-391.
 RP MEDLINE=90001543; PubMed=2506948;
 RA Shima M., Ware J., Yoshioka A., Fukui H., Fulcher C.A.;
 RT "An arginine to cysteine amino acid substitution at a critical
 RT thrombin cleavage site in a dysfunctional factor VIII molecule.";
 RL Blood 74:1612-1617(1989).
 RN [18]
 RN VARIANT HEMA LEU-189.
 RP MEDLINE=90057680; PubMed=2510835;
 RA Chan V., Chan T.K., Tong T.M., Todd D.;
 RT "A novel missense mutation in exon 4 of the factor VIII:C gene
 RT resulting in moderately severe hemophilia A.";
 RL Blood 74:2688-2691(1989).
 RN [19]
 RN VARIANT HEMA LEU-2326.
 RP MEDLINE=89197216; PubMed=2495245;
 RA Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;
 RT "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26
 RT of the factor VIII gene.";
 RL Hum. Genet. 81:335-338(1989).
 RN [20]
 RN VARIANT HEMA HIS-391.
 RP MEDLINE=89264602; PubMed=2498882;
 RA Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Fujimaki M., Hoyer L.W.;
 RT "Direct characterization of factor VIII in plasma: detection of a
 RT mutation altering a thrombin cleavage site
 RT (arginine-372-->histidine).";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
 RN [21]
 RN VARIANT HEMA CYS-1708.
 RP MEDLINE=90105723; PubMed=2104766;
 RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Phillips J.A. III, Janco R.L., Hoyer L.W.;
 RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
 RT Cys) in the factor VIII gene of two unrelated patients with
 RT cross-reacting material-positive hemophilia A.";
 RL Blood 75:384-389(1990).
 RN [22]
 RN VARIANTS HEMA GLN-2228 AND LEU-2326.
 RP MEDLINE=90123183; PubMed=2105106;
 RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,
 RA Mancuso G., Morfini M., de Biasi R., Baudo F., Carbonara A.;
 RT "Recurrent mutations and three novel rearrangements in the factor
 RT VIII gene of hemophilia A patients of Italian descent.";
 RL Blood 75:662-670(1990).
 RN [23]
 RN VARIANT HEMA CYS-391.
 RP MEDLINE=90329422; PubMed=1973901;
 RA Pattinson J.K., McVey J.H., Boon M., Ajani A., Tuddenham E.G.;
 RT "CRM+ haemophilia A due to a missense mutation (372-->Cys) at the
 RT internal heavy chain thrombin cleavage site.";
 RL Br. J. Haematol. 75:73-77(1990).
 RN [24]
 RN VARIANTS HEMA PHE-1699 AND CYS-1708.
 RP MEDLINE=90152691; PubMed=2105906;
 RA Higuchi M., Wong C., Kochan L., Olek K., Aronis S., Kasper C.K.,
 RA Kazazian H.H., Antonarakis S.E.;
 RT "Characterization of mutations in the factor VIII gene by direct
 RT sequencing of amplified genomic DNA.";
 RL Genomics 6:65-71(1990).
 RN [25]
 RN VARIANTS HEMA CYS-1728 AND ASP-1941.
 RP MEDLINE=90169988; PubMed=2106480;
 RA Traystman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,
 RA Kazazian H.H.;
 RT "Use of denaturing gradient gel electrophoresis to detect point
 RT mutations in the factor VIII gene.";

Query Match		94.0%;	Score	7227;	DB	1;	Length	2351;
Best Local Similarity		61.6%;	Pred. No.	0;				
Matches		1437;	Conservative	0;	Mismatches	1;	Indels	894;
		Gaps						
Qy	1	ATRRYYLGAVELSHDYNQSDLGELPLVDARPPRPVKSPFNTSVVYKTLFVBETVHLFN	60					
Db	20	ATRRYYLGAVELSHDYNQSDLGELPLVDARPPRPVKSPFNTSVVYKTLFVBETVHLFN	79					
Qy	61	IAKPRPVMGLGPTTIAQEVYDVTVITLKNASHPVSLHAGVSYKASGAEYDDQTSQ	120					
Db	80	IAKPRPVMGLGPTTIAQEVYDVTVITLKNASHPVSLHAGVSYKASGAEYDDQTSQ	139					
Qy	121	REKEDDVFPFGSGSHYQVQLKENGPMASDPLCLITYSLSHVDLVKDNLSGLIGALLVCR	180					
Db	140	REKEDDVFPFGSGSHYQVQLKENGPMASDPLCLITYSLSHVDLVKDNLSGLIGALLVCR	199					
Qy	181	EGSLAKKTTQTLHFIILLFAVFDGKSWHSTKNSLMQDRDAASARAWPKMHTVNGYVNR	240					
Db	200	EGSLAKKTTQTLHFIILLFAVFDGKSWHSTKNSLMQDRDAASARAWPKMHTVNGYVNR	259					
Qy	241	SLPGLIGCHRKSVMHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFTAOTLL	300					
Db	260	SLPGLIGCHRKSVMHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFTAOTLL	319					
Qy	301	MDLQOFLFCHISSHQHDEMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVR	360					
Db	320	MDLQOFLFCHISSHQHDEMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVR	379					
Qy	361	DDNSPSFIQIRSAKHKPTWVHYIAAEBEDWDYAPLVLPADDRSKYSQVLYNNGPORIG	420					
Db	380	DDNSPSFIQIRSAKHKPTWVHYIAAEBEDWDYAPLVLPADDRSKYSQVLYNNGPORIG	439					
Qy	421	RKYKVRFMAYTDEFTKREAIQHESGILGPLLYGEGVDTLLIIFKQKASRPYNIYPHI	480					
Db	440	RKYKVRFMAYTDEFTKREAIQHESGILGPLLYGEGVDTLLIIFKQKASRPYNIYPHI	499					
Qy	481	TDVRPLYSRRLPKGVKHLKDPFILLPGEIFKFKWTVTVEDGPTKSDPRCLTIYSSFVNME	540					
Db	500	TDVRPLYSRRLPKGVKHLKDPFILLPGEIFKFKWTVTVEDGPTKSDPRCLTIYSSFVNME	559					
Qy	541	RDLASGLIGPLLI CYKESVDQRGQIMSDKENVILFSVFDENRSMWLTENIORFLPNPAG	600					
Db	560	RDLASGLIGPLLI CYKESVDQRGQIMSDKENVILFSVFDENRSMWLTENIORFLPNPAG	619					
Qy	601	VQLEDPFOASNIHMSINGYVFDLSQLSVCLHEVAYWYILSIGAOTDPLSFPSGTYFKH	660					
Db	620	VQLEDPFOASNIHMSINGYVFDLSQLSVCLHEVAYWYILSIGAOTDPLSFPSGTYFKH	679					
Qy	661	KWYEDTTLTLPFGSETVFMSENPGLWILGCHNSDFRNRGWTALLKVSSCDKNTGDYIE	720					
Db	680	KWYEDTTLTLPFGSETVFMSENPGLWILGCHNSDFRNRGWTALLKVSSCDKNTGDYIE	739					
Qy	721	DSYEDISAYLLSKNNALBPRSF-----	742					
Db	740	DSYEDISAYLLSKNNALBPRSFQNSRHPSTRQKQFNATTIPENDIEKTDPFWAHRTMP	799					
Qy	743	-----	742					
Db	800	KIONVSSDLLMLLRQSPTRPHGLSLSDLOEAKYETFSDDPSPGAIDSNNSISEMTHFRPQ	859					
Qy	743	-----	742					
Db	860	LHSGDMVFTPESGQLRLNEKLGTTAATELKKLDFKVSSTSNNLISITIPSDNLAAGTDN	919					
Qy	743	-----	742					
Db	920	TSSLGPPMPVHYDQLDTTLFGKSSPLITESGGPLSLSEENNDKSLLESGLMNSQESSW	979					
Qy	743	-----	742					
Db	980	GKNVSSBESGRLFKGRAHGAPALLTKDNALFKVISILLKTNKTSNNSATNRKTRIDGP	1039					
Qy	743	-----	742					
Db	1040	LIENSPVMQNIILESTEFKVTPLIHDRMLMDKNATALLRLNEMSNKTTSSKNMEMVQOK	1099					
Qy	743	-----	742					
Db	1100	KEGPIPPDAQNPDKSPFKMLFLPESARWQIQRTHGKNSLNSGGQSPKQLVSLGPEKSVEG	1159					
Qy	743	-----	742					
Db	1160	QNFLSEKNKVVVGKEFTKDVGLKEMVFPSSRNFLTNLNDLHNNTHNQEKKIQEBIEK	1219					
Qy	743	-----	742					
Db	1220	KETLIQENVVLPOIHVTGTGTFNFMKNLFLSTRONVEGSDYAYAPVLQDFRSLNDSTNR	1279					
Qy	743	-----	742					
Db	1280	TKKHTAHFSKKGBEENLEGLCNQTKQIVEKVIACCTRIISFNTSQNFVQTSRKALAKQFRL	1339					
Qy	743	-----	742					
Db	1340	PLEBTELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKEKGAIQTQSLDCLTRSHSI	1399					
Qy	743	-----	742					
Db	1400	FOANRSPLPIAKVSSPPSIRPIYLTVLFQDNSSHLPAASYRKDKSGVQESSHFLQAKK	1459					
Qy	743	-----	742					
Db	1460	NNLSAILLTLEMTGDQREVSGSLGTSATNSVYKKVENTVLPKPDLPTXSGKVELLPKVIH	1519					
Qy	743	-----	742					
Db	1520	YQKDLFTETSNPGHLDLVEGSLLOGTEGAIKNNEANRPGKVPFLRVATESAKTPSK	1579					
Qy	743	-----	742					
Db	1580	LLDPLAWNHYGTQIPKEEMKSOBSEKPTAFKKKDTILSINACESNHAIAINEGONKP	1639					
Qy	743	-----SONPVLKXHQREITRTTLOSDQBEIDYDDITISVEMKKEDFDIY	786					
Db	1640	EIEVTWAKQGRTERLCSQNPPLVKRQREITRTTLOSDQBEIDYDDITISVEMKKEDFDIY	1699					
Qy	787	DEDENQSPRSFOKKTIRHYFTAAVERLWDYGMSSSPHVLNRNRAQSGSVPOFKKVVVFQEF	846					
Db	1700	DEDENQSPRSFOKKTIRHYFTAAVERLWDYGMSSSPHVLNRNRAQSGSVPOFKKVVVFQEF	1759					
Qy	847	GSFTQPLYRGELNEHLGLGPIYIRAEVEDNIMVTFRNOASRPYSFYSSLSIYSEEDQROGA	906					
Db	1760	GSFTQPLYRGELNEHLGLGPIYIRAEVEDNIMVTFRNOASRPYSFYSSLSIYSEEDQROGA	1819					
Qy	907	EPRKNFYKPNETKTYFWKVQHHMPTKDEPCKAWAYPSVDLEKDVHSGILGPLVCHT	966					
Db	1820	EPRKNFYKPNETKTYFWKVQHHMPTKDEPCKAWAYPSVDLEKDVHSGILGPLVCHT	1879					
Qy	967	NTLNPAHGRQVTVQEFALFFTIPDETCKSWYFTENNERNCRAPCNIQMEDPTFKENYRPHA	1026					
Db	1880	NTLNPAHGRQVTVQEFALFFTIPDETCKSWYFTENNERNCRAPCNIQMEDPTFKENYRPHA	1939					
Qy	1027	INGYIMDTLPGLVMAQQRIRWYLLSMGSNNIHSIHFSGHVFTVRKKKEKYMALYNLYP	1086					
Db	1940	INGYIMDTLPGLVMAQQRIRWYLLSMGSNNIHSIHFSGHVFTVRKKKEKYMALYNLYP	1999					
Qy	1087	GVFETVEMLSKAGIWRVECLIGELHAGMSTLFLVYSNKCOTPLGMAHGHIROFQITAS	1146					
Db	2000	GVFETVEMLSKAGIWRVECLIGELHAGMSTLFLVYSNKCOTPLGMAHGHIROFQITAS	2059					
Qy	1147	QYQGWAPKLARLHYSINAWSTKEPPSWIKVDLLAPMI IHGKTQAROKFSSLYTSQ	1206					
Db	2060	QYQGWAPKLARLHYSINAWSTKEPPSWIKVDLLAPMI IHGKTQAROKFSSLYTSQ	2119					
Qy	1207	FIIMYSLDGKKQWYRGNSTGTLMVFFGNVDSGKHNIFNPPIIARYIRLHPHTYSIRS	1266					

```

Db 2120 FIIMYSLDKKQWYRGNSTGTLVFFGVNDUSSGKINFIIPPPIARVIRLPHYSIRS 2179
Qy 1267 TIRMELMGCDLNSCMPGMSKAISDAQITASSYFTNNFATWSPSKARLHLQGRSNAWR 1326
Db 2180 TIRMELMGCDLNSCMPGMSKAISDAQITASSYFTNNFATWSPSKARLHLQGRSNAWR 2239
Qy 1327 PQVNNPKWLOVDQKTKMKVGTGVTQGVKSLTSMYKVEFLISSODGHQWTLFQNGKV 1386
Db 2240 PQVNNPKWLOVDQKTKMKVGTGVTQGVKSLTSMYKVEFLISSODGHQWTLFQNGKV 2299
Qy 1387 KVFQGNQDSFPPVNSLDPPLITRYLRHPQSWHQAIRMEVLGCEAQDLY 1438
Db 2300 KVFQGNQDSFPPVNSLDPPLITRYLRHPQSWHQAIRMEVLGCEAQDLY 2351

RESULT 2
FA8_PIG STANDARD; PRT; 2133 AA.
AC P1263; Q95243;
DI 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Coagulation factor VIII precursor (Procoagulant component).
GN F8 OR CF8.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1] _
RP SEQUENCE FROM N.A.
RA Healey J.F., Lubin I.M., Lollar P.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 705-1573 FROM N.A.
RA MEDLINE=86287369; PubMed=3016730;
RX Toole J.J., Pittman D.D., Orr E.C., Murtha P., Wasley L.C.,
RA Kaufman R.J.;
RT "A large region (approximately equal to 95 kDa) of human factor VIII
RT is dispensable for in vitro procoagulant activity.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5939-5942(1986).
RN [3]
RP SEQUENCE OF 392-759 FROM N.A.
RX MEDLINE=94179260; PubMed=7510693;
RA Lubin I.M., Healey J.F., Scandella D., Runge M.S., Lollar P.;
RT "Elimination of a major inhibitor epitope in factor VIII.";
RL J. Biol. Chem. 269:8639-8641(1994).
CC -!- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS
CC AS A COPACITOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE
CC ACTIVATED FORM, FACTOR XA.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENMBL; U49517; AAB06705.1; -.
DR PIR; A25945; A25945.
DR PIR; T42763; T42763.
DR HSSP; P00451; 1CFG.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5 F8 type C; 2.
DR SMART; SM00231; FA58C_2.
DR PROSITE; PS01285; FA58C_1; 2.

```

```

DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00022; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.
KW Blood coagulation; Repeat; Plasma; Acute phase; Calcium;
KW Signal; Glycoprotein; Sulfation.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2133 COAGULATION FACTOR VIII.
FT DOMAIN 20 357 F5/8 TYPE A 1.
FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.
FT DOMAIN 207 357 PLASTOCYANIN-LIKE 2.
FT DOMAIN 399 730 F5/8 TYPE A 2.
FT DOMAIN 399 573 PLASTOCYANIN-LIKE 3.
FT DOMAIN 583 730 PLASTOCYANIN-LIKE 4.
FT DOMAIN 760 1599 F5/8 TYPE A 3.
FT DOMAIN 1495 1822 PLASTOCYANIN-LIKE 5.
FT DOMAIN 1495 1659 PLASTOCYANIN-LIKE 6.
FT DOMAIN 1669 1822 F5/8 TYPE C 1.
FT DOMAIN 1822 1970 F5/8 TYPE C 2.
FT DOMAIN 1975 2127 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 391 392 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 759 760 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1449 1450 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
FT SITE 1490 1491 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT MOD_RES 737 737 SULFATION (BY SIMILARITY).
FT MOD_RES 738 738 SULFATION (BY SIMILARITY).
FT MOD_RES 742 742 SULFATION (BY SIMILARITY).
FT DISULFID 547 573 PROBABLE.
FT DISULFID 1633 1659 PROBABLE.
FT DISULFID 1822 1970 PROBABLE.
FT DISULFID 1975 2127 BY SIMILARITY.
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 601 601 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 929 929 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 985 985 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1025 1025 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1111 1111 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1181 1181 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1208 1208 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1245 1245 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1265 1265 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1335 1335 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1408 1408 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1611 1611 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1919 1919 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 713 713 N -> M (IN REF. 2).
FT CONFLICT 734 734 I -> T (IN REF. 2).
FT CONFLICT 792 792 G -> Q (IN REF. 2).
FT CONFLICT 1133 1133 E -> F (IN REF. 2).
FT CONFLICT 1191 1191 I -> L (IN REF. 2).
FT CONFLICT 1209 1209 R -> F (IN REF. 2).
FT CONFLICT 1437 1437 C -> G (IN REF. 2).
FT CONFLICT 1456 1456 F -> R (IN REF. 2).
FT CONFLICT 1539 1539 F -> R (IN REF. 2).
FT CONFLICT 1546 1546 Q -> N (IN REF. 2).
SQ SEQUENCE 2133 AA; 239304 MW; 152BBAB8997F570DA CRC64;

Query Match 81.0%; Score 6227; DB 1; Length 2133;
Best Local Similarity 57.3%; Pred. No. 0;
Matches 1212; Conservative 106; Mismatches 119; Indels 678; Gaps 3;

Qy 1 ATRRYVLGAVELSDWYMOQSD-LGELPVDARPPRPVKFPFPNTSVYVKKTLFVFTVHLF 59
Db 20 ATRRYVLGAVELSDWYMOQSD-LGELPVDARPPRPVKFPFPNTSVYVKKTLFVFTVHLF 79
Qy 60 NIAKPRPPWMLLGPITQAEVYDVTITLKNASHPVSILHAGVSVYWKASGEAYDDQTS 119
Db 80 SVARPPPPWMLLGPITQAEVYDVTITLKNASHPVSILHAGVSVYWKASGEAYDDQTS 139
Qy 120 QREKEDDKVFPQGSHTYVWVLKENGPMASDPCLCTYTYSVLSHVDLVKDLNSGLIGALLVC 179
Db 140 QREKEDDKVLPKGSQTYVWVLKENGPMASDPCLCTYTYSVLSHVDLVKDLNSGLIGALLVC 199

```


OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX STRAIN=C57BL/6 X CBA; TISSUE=Liver;
 RA MEDLINE=93300511; PubMed=8314577;
 RA Elder B., Iakich D., Gitschier J.;
 RT "Sequence of the murine factor VIII cDNA";
 RL Genomics 16:374-379(1993).
 CC -!- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS
 CC AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE
 CC ACTIVATED FORM, FACTOR XA.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: FOUND IN MOST TISSUES.
 CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
 CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
 CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L05573; AAA37385.1; -;
 DR F1R; A47004; A47004.
 DR HSP; P00451; 1CFG.
 DR MGD; MGI:88383; F8.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR000421; FA58_C.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM0231; FA58C; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 DR PROSITE; PS00222; FA58C_3; 2.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
 DR Blood coagulation; Repeat; Plasma; Acute phase; Calcium;
 KW Signal; Glycoprotein; Sulfation.
 FT SIGNAL 1 19
 FT CHAIN 20 2319
 FT DOMAIN 20 349
 FT DOMAIN 20 199
 FT DOMAIN 207 349
 FT DOMAIN 399 730
 FT DOMAIN 399 573
 FT DOMAIN 583 730
 FT DOMAIN 760 1640
 FT DOMAIN 1683 2008
 FT DOMAIN 1683 1845
 FT DOMAIN 1855 2008
 FT DOMAIN 2008 2156
 FT DOMAIN 2161 2313
 FT SITE 391 392
 FT SITE 759 760
 FT SITE 1678 1679
 FT SITE 1324 1325
 FT SITE 1640 1641
 FT MOD_RES 367 367
 FT MOD_RES 737 737
 FT MOD_RES 738 738
 FT MOD_RES 742 742
 FT MOD_RES 1669 1669
 FT MOD_RES 1687 1687
 FT DISULFID 173 199
 FT DISULFID 547 573
 FT DISULFID 1819 1845
 FT DISULFID 2008 2156
 FT DISULFID 2161 2313
 FT CARBOHYD 61 61
 FT CARBOHYD 233 233

FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 601 601 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 880 880 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 958 958 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1015 1015 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1026 1026 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1076 1076 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1087 1087 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1136 1136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1161 1161 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1192 1192 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1255 1255 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1268 1268 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1273 1273 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1302 1302 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1316 1316 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1340 1340 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1797 1797 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2105 2105 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 2319 AA; 266148 MW; FD054DE051DB2A01 CRC64;
 Query Match 80.5%; Score 6192; DB 1; Length 2319;
 Best Local Similarity 53.4%; Pred. No. 0;
 Matches 1231; Conservative 97; Mismatches 103; Indels 876; Gaps 9;
 QY 1 ATRRYLGAVALSWDYMQSD-LGELPVDAPPRPVPKSPFNTSVVYKTLFVETVHLF 59
 DB 20 ATRRYLGAVALSWYIQSDLLSVLHSDSFLPRMSTSPFNTSIMYKTVFVYKQLF 79
 QY 60 NIAKPRPPWGLLGPTIOAEVYDVTITLKNASHPVSLHAGVSVWKASEGAYDDQTS 119
 DB 80 NIAKPRPPWGLLGPTITWEVHDVTITLKNASHPVSLHAGVSVWKASEGAYDDQTS 139
 QY 120 QREKEDKYPGSHYVWVLKENGPMASDPLCTYSVLSHVDVLDKNSGLIGALLVC 179
 DB 140 QMEKDDKYPGESHTYVWQVLKENGPMASDPLCTYSVLSHVDVLDKNSGLIGALLVC 199
 QY 180 REGSLAKEKTQTLHKFILLFAVDFECKSWHSETKNSLMQDRDAASARAWPKHVTNGYN 239
 DB 200 KEGSLKERTOMLYQFVLLFAVDFECKSWHSEKNDSTYQMSDSASARDWPKHVTNGYN 259
 QY 240 RSLPLGICHRKSVYVHWITGMGTPPEVHSIFLEHTFLVRNHRQASLEISPTFLTAQTL 299
 DB 260 RSLPLGICHRKSVYVHWITGMGTPPEVHSIFLEHTFLVRNHRQASLEISPTFLTAQTL 319
 QY 300 LMDLGQFLAFCHISSHOHGMEAYVKVDCPEPQRLMK-NNEEAEDYDDDLTDSEMDVV 358
 DB 320 LIDLQGLFCHISSHKHGMEAYVKVDCPEESQWQKNNNEEMEDYDDDLTDSEMDVF 378
 QY 359 RPDNDSPSFQIRSVAKHKPKTWVHYIAAEEDWDYAPLVAPDERSYKSQYLNNGPQR 418
 DB 379 TLDYDSSP-FIQIRSVAKKPKTWIHYIAAEEDWDYAPSVPTSDNGSYKSQYLSNGPQR 437
 QY 419 IGRYKVKRPMATDETFTKTRAIQHESGLIGPLLYGEVGDITLLIFKQASRPYNIYPH 478
 DB 438 IGRYKVKRPIAYTDETFTKTRAIQHESGLIGPLLYGEVGDITLLIFKQASRPYNIYPH 497
 QY 479 GITDVPYLSRRLPKGVKHLKOPPLIPGRIKPKTWVTVDEGPTKSDPRCLTYYSFVN 538
 DB 498 GITDVPYLSRRLPKGVKHLKOPPLIPGRIKPKTWVTVDEGPTKSDPRCLTYYSFVN 557
 QY 539 MERDLASGLIGPLLYCYKESVDQGRNQIMSDKXNVILFSVFDENRSMYLTENIQRLPNP 598
 DB 558 PERDLASGLIGPLLYCYKESVDQGRNQIMSDKXNVILFSIFDENQWYITENQRLPNA 617
 QY 599 AGVQLEDPEFQASNMHSINGYVFDLSQLSVCLHEVAYVYLLSIGAQDTFLSFFGYTF 658
 DB 618 AKTQPDGPGQASNMHSINGYVFDLSQLSVCLHEVAYVYLLSIGAQDTFLSFFGYTF 677

RT of internal repeats.";
 RL Biochemistry 26:6508-6514 (1987).
 RN [4]
 RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
 RX MEDLINE=86313665; PubMed=3092220;
 RA Kane W.H., Davie E.W.;
 RT "Cloning of a cDNA coding for human factor V, a blood coagulation
 factor homologous to factor VIII and ceruloplasmin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804 (1986).
 RN [5]
 RP PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=93203619; PubMed=8454869;
 RA Shen N.L.L., Fan S.-T., Pyati J., Graff R., Lapolla R.J.,
 Edgington T.S.;
 RT "The serine protease cofactor factor V is synthesized by
 lymphocytes.";
 RL J. Immunol. 150:2992-3001 (1993).
 RN [6]
 RP SULFATION.
 RX MEDLINE=94264012; PubMed=8204629;
 RA Pittman D.D., Tomkinson K.N., Michnick D., Selighsohn U.,
 Kaufman R.J.;
 RT "Posttranslational sulfation of factor V is required for efficient
 thrombin cleavage and activation and for full procoagulant activity.";
 RL Biochemistry 33:6952-6959 (1994).
 RN [7]
 RP SULFATION.
 RX MEDLINE=90366699; PubMed=2168225;
 RA Hortin G.L.;
 RT "Sulfation of tyrosine residues in coagulation factor V.";
 RL Blood 76:946-952 (1990).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
 RX MEDLINE=20052169; PubMed=10586886;
 RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
 Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
 Fuentes-Prior P.;
 RT "Crystal structures of the membrane-binding C2 domain of human
 coagulation factor V.";
 RL Nature 402:434-439 (1999).
 RN [9]
 RP VARIANT MET-1764.
 RX MEDLINE=95179146; PubMed=7874144;
 RA Bayston T.A., Ireland H., Olds R.J., Thein S.L., Lane D.A.;
 RT "A polymorphism in the human coagulation factor V gene.";
 RL Hum. Mol. Genet. 3:2085-2085 (1994).
 RN [10]
 RP VARIANT APCR GLN-534.
 RX MEDLINE=94217810; PubMed=8164741;
 RA Bertina R.M., Koelmen B.P.C., Koster T., Rosendaal F.R.,
 Dirven R.J., de Ronde H., van der Velden P.A., Reitsma P.H.;
 RT "Mutation in blood coagulation factor V associated with resistance to
 activated protein C.";
 RL Nature 369:64-67 (1994).
 RN [11]
 RP VARIANTS ILE-1285 AND ARG-1327.
 RX MEDLINE=96351768; PubMed=8713778;
 RA Lunghi B., Iacoviello L., Gemmati D., Dilasio M.G., Castoldi E.,
 Pinotti M., Castaman G., Redaelli R., Mariani G., Marchetti G.,
 Bernardi F.;
 RT "Detection of new polymorphic markers in the factor V gene:
 association with factor V levels in plasma.";
 RL Thromb. Haemost. 75:45-48 (1996).
 RN [12]
 RP VARIANT APCR GLY-334, AND VARIANT LYS-513.
 RX MEDLINE=98122763; PubMed=9454741;
 RA Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;
 RT "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese.";
 RL Blood 91:1135-1139 (1998).
 RN [13]
 RP VARIANT APCR THR-334.
 RX MEDLINE=98122764; PubMed=9454742;

RA Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;
 RT "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
 resistance to activated protein C.";
 RL Blood 91:1140-1144 (1998).
 RN [14]
 RP VARIANTS HIS-107; THR-413; LYS-513; SER-809; THR-817; ARG-858;
 RX ARG-865; GLU-925; GLN-1146; ALA-1530; SER-1685; VAL-1749; MET-1764;
 RP ILE-1820 AND GLY-2222, AND VARIANT APCR GLN-534.
 RX MEDLINE=99318093; PubMed=10391209;
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
 Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nemesh J., Ziaugra L.,
 Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
 Lander E.S.;
 RA "Characterization of single-nucleotide polymorphisms in coding regions
 of human genes.";
 RT Nat. Genet. 22:231-238 (1999).
 RL [15]
 RP ERRATUM.
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
 Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nemesh J., Ziaugra L.,
 Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
 Lander E.S.;
 RA Nat. Genet. 23:373-373 (1999).
 RL CC -!- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
 WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
 CC CC -!- SURUNIT: Factor Va is composed of a heavy chain and a light
 chain, noncovalently bound. The interaction between the two chains
 is calcium-dependent.
 CC CC -!- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
 REPEATS.
 CC CC -!- PTM: Thrombin activates factor V proteolytically to the active
 cofactor, factor Va (formation of a heavy chain at the N-
 terminus and a light chain at the C-terminus).
 CC CC -!- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
 ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY.
 CC CC -!- DISEASE: Defects in F5 are the cause of Owren parahemophilia
 [MIM:227400], an hemorrhagic diathesis.
 CC CC -!- DISEASE: Defects in F5 are the cause of resistance to activated
 protein C (APCR) [MIM:188055], a form of thrombophilia. The APCR
 mutation is found in about 5% of the population which suggest that
 a slight thrombotic tendency may confer some advantage in fetal
 implantation.
 CC CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
 CC CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
 CC CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
 CC CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L32779; AAB59401.1; JOINED.
 DR EMBL; L32755; AAB59401.1; JOINED.
 DR EMBL; L32756; AAB59401.1; JOINED.
 DR EMBL; L32757; AAB59401.1; JOINED.
 DR EMBL; L32758; AAB59401.1; JOINED.
 DR EMBL; L32759; AAB59401.1; JOINED.
 DR EMBL; L32760; AAB59401.1; JOINED.
 DR EMBL; L32761; AAB59401.1; JOINED.
 DR EMBL; L32762; AAB59401.1; JOINED.
 DR EMBL; L32763; AAB59401.1; JOINED.
 DR EMBL; L32764; AAB59401.1; JOINED.
 DR EMBL; L32765; AAB59401.1; JOINED.
 DR EMBL; L32766; AAB59401.1; JOINED.
 DR EMBL; L32767; AAB59401.1; JOINED.
 DR EMBL; L32768; AAB59401.1; JOINED.
 DR EMBL; L32769; AAB59401.1; JOINED.
 DR EMBL; L32770; AAB59401.1; JOINED.
 DR EMBL; L32771; AAB59401.1; JOINED.
 DR EMBL; L32772; AAB59401.1; JOINED.

FT	DOMAIN	1748	1890	PLASTOCYANIN-LIKE 6.	
FT	DOMAIN	1894	2048	F5/8 TYPE C 1.	
FT	DOMAIN	2053	2208	F5/8 TYPE C 2.	
FT	SITE	741	742	CLEAVAGE (BY THROMBIN)	(BY SIMILARITY)
FT	SITE	1034	1035	CLEAVAGE (BY THROMBIN)	(BY SIMILARITY)
FT	SITE	1564	1565	CLEAVAGE (BY THROMBIN)	(BY SIMILARITY)
FT	DISULFID	167	193	PROBABLE.	
FT	DISULFID	499	525	PROBABLE.	
FT	DISULFID	1712	1738	PROBABLE.	
FT	DISULFID	1894	2048	BY SIMILARITY.	
FT	DISULFID	2053	2208	BY SIMILARITY.	
FT	MOD RES	697	697	SULFATION (POTENTIAL).	
FT	MOD RES	701	701	SULFATION (POTENTIAL).	
FT	MOD RES	730	730	SULFATION (POTENTIAL).	
FT	MOD RES	1513	1513	SULFATION (POTENTIAL).	
FT	MOD RES	1529	1529	SULFATION (POTENTIAL).	
FT	MOD RES	1537	1537	SULFATION (POTENTIAL).	
FT	MOD RES	1541	1541	SULFATION (POTENTIAL).	
FT	CARBOHYD	225	225	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	297	297	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	382	382	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	460	460	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	553	553	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	587	587	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	745	745	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	756	756	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	774	774	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	780	780	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	902	902	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	952	952	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	964	964	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1044	1044	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1053	1053	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1062	1062	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1071	1071	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1078	1078	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1094	1094	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1451	1451	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1490	1490	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1550	1550	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1690	1690	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1839	1839	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1997	1997	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2196	2196	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	VARIANT	587	592	NFTAPA -> T (IN VARIANT 2).	
SQ	SEQUENCE	2211	AA; 248981	MM; CBBF90B738667C45	CRC64;
Query Match					30.8%; Score 2372; DB 1; Length 2211;
Best Local Similarity					26.2%; Pred. No. 8.8e-147;
Matches					594; Conservative 265; Mismatches 487; Indels 924; Gaps 35;
Qy	3	RRYLGAVELSWDYNQSDLGELPVDARPPRPVPSFPNTSVVYKTLFVETVHLFNIA	62		
Db	32	RQFYAAQSIRWNYR-----PESTHL-----SSKEPETS--PKIVIREYRAY-FQKE	76		
Qy	63	KRPPEWGLLGTTOAEVDTWTLTKMASHPVSLHAGVSYWKASGEAYDDOTSORE	122		
Db	77	KPQRTSGLLGTTHYAEVDTWTLTKMASHPVSLHAGVSYWKASGEAYDDOTSORE	136		
Qy	123	KEDDKVPFGGSHYTVWQVLKENGPMASDPLCLITYSLSHVDLVKDLNSGLICALLVCREG	182		
Db	137	KMDAVAPQBYTYEYIIEHSGPHDDPPCLTHIYYSVNLVEDFNSGLIGPLICKG	196		
Qy	183	SLAKEKTTL--HKFILLFVDFDEKSWHSEKNSLMODRDAASARAWPKMHTVNGVYNR	240		
Db	197	TLTSDGTOKMFEKQHVLMFAVEDSKSNQTS-----LMYTVNGVYNG	240		
Qy	241	SLPLGLICHRKSVVWHVIGMTYEVHSIFLEGHTFLVYNRQASLEISPIFTLTAQLL	300		
Db	241	TMPDITVCAHDHISWHLGSSGPELFSIFHNGQVLEQNHKISAITLVSATSTTANMTV	300		
Qy	301	MDLQGLFLFCHISSHQHDGMEAYVKVDSCEPPEQLRMKNNEAEYDDDLTSDMDVYRF	360		

Db	301	SPEGRWTIASLPRHQAGQAYIDHKCAKTRNPKK-----LTRDQ-----	343		
Qy	361	DDNSPSFIQIRS VAKGPKTKWVHYIAAEEDMDYAPLVLPADDRSKYKSOVNNPGORIG	420		
Db	344	-----RRHKKWEYFIAAEVINDYAPIIPANMDKKTRSLHLDNFSNRIG	388		
Qy	421	RKYKVRPMAYTDTFKTR--EAIQHESGILGPLLYGEVGDTLIIIFKNOASREYNYIYPH	478		
Db	389	KHYKVVYKQYQDDSFTRKLEDPSGEGILGPIIRAQVRDTLKIVFNKMASRSYSIVPH	448		
Qy	479	GIT-----DVRPLYSRRLPKGVKHLKDFPILPGEIPKYYKWTVTVEGPTKSDRCL	529		
Db	449	GVTFSPYDNEVNSSTSGSNTMIRAVR-----PGETTYKWNILSEDEPTENDACQL	500		
Qy	530	TRYSSVFNMERDLASGLIGLILICYKESVDORGNQIMSKRNVLFSVPDENRSWYLTE	589		
Db	501	TRPYSNVDITRDLASGLIGLILICKERSIDRRGIQRAADIEQQAFAVDENKSWIIE	560		
Qy	590	NIQRFLPNPAGVQLEDPEFOASNMH-----SINGYVFDLSQ--LSVCILHEVAYWYILSIG	643		
Db	561	NIYKFCENPEKVKRDDPKFYESNIMSNFTLPAINGYVPESIPILGFCFDDTVQWHPCSVG	620		
Qy	644	AQTDPLSVFSGYTFKHKMYVEDTLTLFPFSGETVFMENPGLWILGCTHNSDPRNGMT	703		
Db	621	TONDILTIHFTGHSFYIKRHEDTLTLFPMQGESVTVMNVGTWMLTWNNSPRSKKL	680		
Qy	704	ALLKVSCKDKTGD-----YYEDS-----YEDISAYLL-----	731		
Db	681	LRFDACKIRNDDDDSYEIIYFSGSTAMTKYKHDSIEDENDADSDYQDELALILGL	740		
Qy	732	-----SKNNAISPRFSQ-----	744		
Db	741	RSPFNSSINQKDELNTALALEKDSFIPPSANRSLDSNSSSRSHVSRUJAKNFBASLX	800		
Qy	745	-----NPPVLK-----	750		
Db	801	TLHLLEAPAGSPLEHAGLKGKNGSALNPPMAEHSPPSYSEDPRDHPLSDVTGVSLLPFGT	860		
Qy	751	-----RHOR-----EYTRTTLQSDQ-----EEI-----	768		
Db	861	FKNRKPAKHORFQVGRGQAQKHFSQTRFPAKTRTSLSDQNSSSRMGWEDIPSDLL	920		
Qy	769	-----	768		
Db	921	LQKQDPYKILNGEWHLYSEKSYEIIQDANENKTVNKLPSNPQSDSRWTGENIFPKNSHG	980		
Qy	769	-----	768		
Db	981	KQSGHPTFLVTRRKPLQDRDRNRSLKEGLPLIRTRKKKEKPAYHVPLSPRSFHLR	1040		
Qy	769	-----DYDD-----	772		
Db	1041	GEVNASFDRRRHNSLLHLLHASNETSIDLNQTFPSMNLASLASLPDHDQSPNDTTSQT	1100		
Qy	773	-----TISVENKMKDFDIYDED-----ENQSP-----	794		
Db	1101	SSPDLTYTSPPEHYQIFPIQSDPHTHTTAPSNRSPDPHTHTTAPSNRSPDPHTTAP	1160		
Qy	795	-----	794		
Db	1161	NYDLNRRAIPTDVSOIIPFSLVLEWQATATSLDLSQPSISPLDQWALLSPDQESLSPDL	1220		
Qy	795	-----	794		
Db	1221	QOTSLSPDLQESLSPDLQOTALSPPDPSQESLSPDLQOTALSPPDPSQESLSPDLQOTALS	1280		
Qy	795	-----	794		
Db	1281	PDPQESLSPDLQOTALSPPDPSQESLSPDLQOTALSPPDPSQESLSPDLQOTALSPPDPSQ	1340		
Qy	795	-----	794		

Db 1341 SLSPDLGQTSPLDQBSLSPLDGTALSPDPQBSLSPLDQTSPLDQBSLSPLD 1400
 QY 795 ----- 794
 Db 1401 GQTALSPDLQBSLSPLDQTSPLDQBSLSPLDQTSPLDQBSLSPLDQTSPLD 1460
 QY 795 ----- 794
 Db 1461 QSLPLPFGQTFNADIGQSPDPSTLNTFIPDFNPLVVGSLRDDGYIELIPRQ 1520
 QY 795 ----- RSFQKTRHYFAAVER 811
 Db 1521 KEESSEEDYGEFVAYNDPYQTLRTDINSRNPDAWYLRGNTGRKYVIAAEI 1580
 QY 812 LNDYGMSSPHVLRNRAQSGS---VPO---FKVYFQBTGSGFTQPLRYGELNHLGL 865
 Db 1581 SWDYS-----RFVQSDDDYVPEFTVYKVVFRKYLDTFTKLPDQGEYEHGL 1631
 QY 866 GPVIRAEVDNIMVFRNQAASRPYSFYSLISYE-----EDQRQGAEPKRNFKVNE 917
 Db 1632 GPVIRAEVDNIMVFRNQAASRPYSFYSLISYE-----EDQRQGAEPKRNFKVNE 1691
 QY 918 TKTYFWKVOHMAPTKDEFCWAYFSDVDLEKDVHSLGLGLPLVCHTNTLNPAHGRQV 977
 Db 1692 TVTYVWHATTRSGPENPGSACRWAYSAVNPKEKDIHSLGLGLPLVCHTNTLNPAHGRQV 1751
 QY 978 TVQEFALFTTIDETKSWYFTENMRNCRAPCNQIOWEDTFFKYNRFAHNGIYMDTLPG 1037
 Db 1752 DMREFVLLFMVDEKSWYDKKPTRSRWRASS-----EVKNSHEFAHNGIYMDTLPG 1804
 QY 1038 LVMAQDQIRWLLSGNSNENIHSIFSHGVFTVRKKEKYNALNLYPGVETVEMLPS 1097
 Db 1805 LEMVQEWVRLHLLNGSRDILHVHFGHQTILENGTQHQHGVWPLPGSKYLEMKAS 1864
 QY 1098 KAGIWRVECLIGEHLHAGMSTLFLVYNNKQPTPLGWSAGHIRDFQITASQYQWAPKLA 1157
 Db 1865 KPGWLLDTEVGEIQAGNQTPFLIVDRECKPMGLSTGLIADSQIQASEFNGYWEFKLA 1924
 QY 1158 RLHYSGSINAW-----STK-EFWSIKVDLLAPMIHIGTKTGAKQKSSLYISQIFIMY 1211
 Db 1925 RLNNGSGYNAMTAELSTEFNPEPIQVDMQKVELLGTQAGKHYLKPYYTTFECVAY 1984
 QY 1212 SLDGKWKQYRNGSTGLTVFVNGVDSSGKINENIPPIIARYIKLHPHYISRTLRME 1271
 Db 1985 SLDRKNWRLFKGNSTNVMVFGNSDASTIKENQIDPPVARYIISPTGSKYKALRLE 2044
 QY 1272 LMGCDLNCSMPLGMSKASDAQITASSYFTNMFAT-WSPSKARHLHQGRNARPOVN 1330
 Db 2045 LQCEVNGCGSTFLGMESGKIENKQITASSFKKSWGNYWEPLARLNAQGRVNAQAKAN 2104
 QY 1331 NPKWLVQDFQTKMKTGVTGTQGVKSLTSMVYKBEFLISSQDGHQWTLFFQNGKV--KV 1388
 Db 2105 NNQWLVQDFQTKMKTGVTGTQGVKSLTSMVYKBEFLISSQDGHQWTLFFQNGKV--KV 2164
 QY 1389 FQGNODSPVNVSLDPLLTLYRLTHPQSWHQIALRMEVLGCEAQDLY 1438
 Db 2165 FEGNNVGRHVKNFNPPIISIRFIRIIPKTNQSIARLELPGC----DMY 2211

RESULT 6

FA5_PIG STANDARD; PRT; 2258 AA.
 AC O9GLE1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Coagulation factor V precursor (Activated protein C cofactor).
 GN F5.
 OS Sus. scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]

RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C
 RP DOMAINS.
 RC TISSUE=Liver;
 RX MEDLINE=21121490; PubMed=11229814;
 RA Grimm D.R., Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,
 RA Kim H.K.W.;
 RT "Porcine factor V: cDNA cloning, gene mapping, three-dimensional
 RT protein modeling of membrane binding sites and comparative anatomy of
 RT domains." Life Sci. 58:148-159(2001).
 RL CC -!- FUNCTION: Coagulation factor V is a cofactor that participates
 CC with factor Xa to activate prothrombin to thrombin.
 CC -!- SUBUNIT: Factor Va is composed of a heavy chain and a light
 CC chain, noncovalently bound. The interaction between the two chains
 CC is calcium-dependent.
 CC -!- DOMAIN: Domain B contains 41 X 9 AA tandem repeats. Domains C1
 CC and C2 may be involved in membrane binding.
 CC -!- PTM: Thrombin activates factor V proteolytically to the active
 CC cofactor, factor Va (formation of a heavy chain at the N-
 CC terminus and a light chain at the C-terminus).
 CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
 CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
 CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF191308; AAG28381.1; --
 DR HSPF; P12259; 1CZT.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR000421; FA58_C.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM00231; FA58C_2; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 DR PROSITE; PS00022; FA58C_3; 2.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
 DR Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
 DR Repeat.
 KW SIGNAL. 1 22
 KW CHAIN. 23 2258
 KW CHAIN. 23 737
 KW PEPTIDE. 738 1611
 FT CHAIN. 1612 2258
 FT DOMAIN. 30 329
 FT DOMAIN. 30 193
 FT DOMAIN. 203 329
 FT DOMAIN. 348 683
 FT DOMAIN. 348 525
 FT DOMAIN. 535 683
 FT DOMAIN. 691 1611
 FT DOMAIN. 1168 1539
 FT REPEAT. 1168 1176
 FT REPEAT. 1177 1185
 FT REPEAT. 1186 1194
 FT REPEAT. 1195 1203
 FT REPEAT. 1204 1212
 FT REPEAT. 1213 1221
 FT REPEAT. 1222 1230
 FT REPEAT. 1231 1239
 FT REPEAT. 1240 1248
 FT REPEAT. 1249 1257
 FT REPEAT. 1258 1266
 FT REPEAT. 1267 1275
 FT REPEAT. 1276 1284

41 X 9 AA APPROXIMATE TANDEM REPEATS OF
 T-L-S-P-D-L-[GS]-[HQ]-T.

RT "Single-chain structure of human ceruloplasmin: the complete amino-
acid sequence of the whole molecule.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:390-394(1984).
RN [6]
RP SEQUENCE OF 158-333; 518-724 AND 858-1065.
RX MEDLINE=83117800; PubMed=6571985;
RA Takahashi N., Bauman R.A., Otel T.L., Dwulet F.E., Wang C.-C.,
RN Putnam F.W.;
RT "Internal triplication in the structure of human ceruloplasmin.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:115-119(1983).
RN [7]
RP SEQUENCE OF 501-905.
RX MEDLINE=81199407; PubMed=6940148;
RA Dwulet F.E., Putnam F.W.;
RN "Complete amino acid sequence of a 50,000-dalton fragment of human
ceruloplasmin.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:790-794(1981).
RN [8]
RP SEQUENCE OF 907-1065.
RX MEDLINE=80137543; PubMed=6987229;
RA Kingston I.B., Kingston B.L., Putnam F.W.;
RN "Primary structure of a histidine-rich proteolytic fragment of human
ceruloplasmin. I. Amino acid sequence of the cyanogen bromide
peptides.";
RL J. Biol. Chem. 255:2878-2885(1980).
RN [9]
RP SEQUENCE OF 907-1065.
RX MEDLINE=80137544; PubMed=6987230;
RA Kingston I.B., Kingston B.L., Putnam F.W.;
RN "Primary structure of a histidine-rich proteolytic fragment of human
ceruloplasmin. II. Amino acid sequence of the tryptic peptides.";
RL J. Biol. Chem. 255:2886-2896(1980).
RN [10]
RP SEQUENCE OF 1007-1061 FROM N.A.
RX MEDLINE=90285218; PubMed=2355023;
RA Yang F.M., Friedrichs W.E., Cupples R.L., Banifacio M.J.,
RN Sanford J.A., Horton W.A., Bowman B.H.;
RT "Human ceruloplasmin. Tissue-specific expression of transcripts
produced by alternative splicing.";
RL J. Biol. Chem. 265:10780-10785(1990).
RN [11]
RP REVIEW.
RX MEDLINE=22049919; PubMed=12055353;
RA Hellman N.E., Gitlin J.D.;
RN "Ceruloplasmin metabolism and function.";
RL Annu. Rev. Nutr. 22:439-458(2002).
RN [12]
RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS).
RA Zaitseva I., Zaitsev V., Card G., Moshkov K., Bax B., Ralph A.,
RN Lindley P.;
RT "The X-ray structure of human serum ceruloplasmin at 3.1 A: nature of
the copper centres.";
RL J. Biol. Inorg. Chem. 1:15-23(1996).
CC -1- FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING [6-7 ATOMS PER
MOLECULE] GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS
ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT
AND HOMEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
CC -1- COFACTOR: BINDS 6 CU-IONS PER MOLECULE. THIS PROTEIN BELONGS TO
THE MULTICOPPER OXIDASES WHICH CONTAIN THREE DISTINCT CU CENTERS
KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE 3 OR COUPLED
BINUCLEAR.
CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN LIVER AND SECRETED INTO THE
PLASMA.
CC -1- DISEASE: Defects in CP are the cause of aceruloplasminemia
[MIM:604290], an autosomal recessive disorder of iron metabolism.
CC It is characterized by iron accumulation in the brain as well as
visceral organs. Clinical features consist of the triad of retinal
degeneration, diabetes mellitus and neurological disturbances.
CC -1- DISEASE: Ceruloplasmin levels are decreased in Wilson's disease,
in which copper cannot be incorporated into ceruloplasmin in
liver because of defects in the copper-transporting ATPase 2.
CC -1- SIMILARITY: Contains 3 F5/8 type A domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M13699; AAA51976.1; --
DR EMBL; D45045; BAA08085.1; --
DR EMBL; D45044; BAA08084.1; --
DR EMBL; D45028; BAA08084.1; JOINED.
DR EMBL; D45029; BAA08084.1; JOINED.
DR EMBL; D45030; BAA08084.1; JOINED.
DR EMBL; D45031; BAA08084.1; JOINED.
DR EMBL; D45032; BAA08084.1; JOINED.
DR EMBL; D45033; BAA08084.1; JOINED.
DR EMBL; D45034; BAA08084.1; JOINED.
DR EMBL; D45035; BAA08084.1; JOINED.
DR EMBL; D45036; BAA08084.1; JOINED.
DR EMBL; D45037; BAA08084.1; JOINED.
DR EMBL; D45038; BAA08084.1; JOINED.
DR EMBL; D45039; BAA08084.1; JOINED.
DR EMBL; D45040; BAA08084.1; JOINED.
DR EMBL; D45041; BAA08084.1; JOINED.
DR EMBL; D45042; BAA08084.1; JOINED.
DR EMBL; D45043; BAA08084.1; JOINED.
DR EMBL; D00025; BAA00019.1; --
DR EMBL; X04135; CAA27752.1; --
DR EMBL; X04136; CAA27753.1; --
DR EMBL; X04137; CAA27754.1; --
DR EMBL; X04138; CAA27755.1; --
DR EMBL; M13536; AAA51975.1; --
DR EMBL; J05506; --; NOT_ANNOTATED_CDS.
DR PIR; A25443; KUHU.
DR PDB; 1KCW; 12-FEB-97.
DR GlycoSuiteDB; P00450; --
DR SWISS-2DPAGE; P00450; HUMAN.
DR Slena-2DPAGE; P00450; --
DR Genew; HGNC:2295; Cp.
DR MIM; 117700; --
DR MIM; 604290; --
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0006879; P:iron ion homeostasis; TAS.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR002355; MultiCu-oxidase2.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER OXIDASE2; 1.
DR Oxidoreductase; Copper; Metal-binding; Glycoprotein; Plasma; Repeat;
KW Signal; Polymorphism; 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 1065 CERULOPLASMIN.
FT F5/8 TYPE A 1.
FT DOMAIN 20 357 PLASTOCYANIN-LIKE 1.
FT DOMAIN 20 200 PLASTOCYANIN-LIKE 2.
FT DOMAIN 209 357 PLASTOCYANIN-LIKE 2.
FT DOMAIN 370 718 F5/8 TYPE A 2.
FT DOMAIN 370 560 PLASTOCYANIN-LIKE 3.
FT DOMAIN 570 718 PLASTOCYANIN-LIKE 4.
FT DOMAIN 730 1061 F5/8 TYPE A 3.
FT DOMAIN 730 900 PLASTOCYANIN-LIKE 5.
FT DOMAIN 908 1061 PLASTOCYANIN-LIKE 6.
FT CARBOHYD 138 136 N-LINKED (GLCNAC. . .).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .).
FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .).
FT CARBOHYD 762 762 N-LINKED (GLCNAC. . .).
FT DISULFID 174 200 PROBABLE.
FT DISULFID 276 357 PROBABLE.
FT DISULFID 534 560 PROBABLE.
FT DISULFID 637 718 PROBABLE.
FT DISULFID 874 900 PROBABLE.
FT METAL 120 120 COPPER (TYPE 2) (BY SIMILARITY).

DR HSP; P00740; 1EDM.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR006210; IEGF.
DR Pfam; PF000008; EGF; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00181; EGF_2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00022; FA58C_3; 2.
KW Glycoprotein; Repeat; EGF-like domain.
FT DOMAIN 2 41 EGF-LIKE 1.
FT DOMAIN 44 88 EGF-LIKE 2.
FT DOMAIN 91 247 F5/8 TYPE C 1.
FT DOMAIN 252 409 F5/8 TYPE C 2.
FT SITE 67 69 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 6 17 BY SIMILARITY.
FT DISULFID 11 29 BY SIMILARITY.
FT DISULFID 31 40 BY SIMILARITY.
FT DISULFID 91 247 BY SIMILARITY.
FT DISULFID 234 238 BY SIMILARITY.
FT DISULFID 252 409 BY SIMILARITY.
FT CARBOHYD 41 41 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 409 AA; 45725 MW; BOC07AF80029927A CRC64;

Query Match 8.5%; Score 650; DB 1; Length 409;
Best Local Similarity 39.3%; Pred. No. 3.7e-35;
Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

QY 1105 ECLIGELHAGMSTFLVYSNK-----COTPLGWSAGHRIHQITASGOY-- 1149
DB 58 ECEVIDAHRG--DVFEYICKPHGYTGTHCEIICNAPMETGATFOISSAHLG 115
QY 1150 ----GOWAPKLARLHYSGSINAW--STKEPFSWIKVDLLAPMIITHGKTQCAROKFSSLY 1203
DB 116 FMGLQWAPELARLHRAIGVNAWASNDYRNPMIQVNLRRMRVTGVVTOGASPAAGAEY 175
QY 1204 ISQFIWYSLDGKKWQYRNGSTGLMVFFGNVDSSGKINENPPIIARYIRLHPHYS 1263
DB 176 MKTFKVAYSTGRKFQFQIOGAESGDKIFMGNLNSGLKVNLFEPVLEVQYVRLVPIICH 235
QY 1264 IRSTRLMELMGCDLNSCMPLGMSKASDAQITASSYFTN---MPATWSPSKARLHLOG 1320
DB 236 RGCTLAPELLGCELSCGAEPGLKNDTIPKQITASSFYRTWGLSFSWTFYARLDNQG 295
QY 1321 RSNARPPQVNPKEWLQVDFQKTKMVTGVTQGVKSLTSMYKFLISSQDGHQWTLF 1380
DB 296 KFNAWTAQNSASEWLQIDLGSQRRTGIIITQAGDFGHQYVAAVKVAYDDGVSWTEY 355
QY 1381 FQNGKV--KVFQGNQDSFTPVNSLDPLTLRYLRHPQSWVHQIALRMEVLGC 1432
DB 356 RDQGALEGKIFPGNLDNSHKKNMFETPLTRFVRLPVAHNRIILRVLLGC 409

RESULT 13

MFGM_BOVIN STANDARD; PRT; 427 AA.
AC Q95114; P79344; Q27959;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)
DE (MGP57/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein
DE SP47) (BP47) (Components 15/16).
GN MFG8.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;
RN [3]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RX STRAIN=Holstein; TISSUE=Mammary gland;
RX MEDLINE=97008954; PubMed=8856064;
RA Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
RA Petersen T.B.;
RT "Characterization of glycoprotein PAS-6/7 from membranes of bovine
RT milk fat globules.";
RL Eur. J. Biochem. 240:628-636 (1996).
RN [2]
RP SEQUENCE OF 18-427 FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=96125736; PubMed=8541316;
RA Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
RA Mameda T.;
RT "Molecular cloning of glycoprotein antigens MGP57/53 recognized by
RT monoclonal antibodies raised against bovine milk fat globule
RT membrane.";
RL Biochim. Biophys. Acta 1245:385-391 (1995).
RN [3]
RP SEQUENCE OF 19-427 FROM N.A.
RC TISSUE=Testis;
RX Russlin M.A.;
RT Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
RC TISSUE=Milk;
RX MEDLINE=93250576; PubMed=8485470;
RA Mather I.H., Banghart L.R., Lane W.S.;
RT "The major fat-globule membrane proteins, bovine components 15/16 and
RT guinea-pig GP 55, are homologous to MGF-E8, a murine glycoprotein
RT containing epidermal growth factor-like and factor V/VIII-like
RT sequences.";
RL Biochem. Mol. Biol. Int. 29:545-554 (1993).
CC -1- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
CC MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
CC BINDING PROTEIN.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q95114-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q95114-2; Sequence=VSP_001398;
CC -1- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN
CC -1- PTM: THE 2 O-LINKED GLYCANS CONSTITUTE OF GAL, GLCNAC AND FUC, WITH
CC PROBABLY FUC AS REDUCING TERMINAL SUGAR.
CC -1- SIMILARITY: Contains 2 EGF-like domains.
CC -1- SIMILARITY: Contains 2 F5/8 type C domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X91895; CAA62997.1; -;
DR EMBL; S80643; AAB35894.2; -;
DR EMBL; Y11719; CAA72406.1; -;
DR PIR; S74211; S74211.
DR HSP; P00740; 1IXA.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR006210; IEGF.
DR Pfam; PF000008; EGF; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00181; EGF_2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01285; FA58C_1; 2.

```
DR PROSITE; PS01286; FA58C.2; 2.  
DR PROSITE; PS00222; FA58C.3; 2.  
KW Signal; Glycoprotein; Milk; Repeat; EGF-like domain;  
FT SIGNAL 1 18  
FT CHAIN 19 427 LACTADHERIN.  
FT DOMAIN 20 59 EGF-LIKE 1.  
FT DOMAIN 62 106 EGF-LIKE 2.  
FT DOMAIN 109 265 F5/8 TYPE C 1.  
FT DOMAIN 270 427 F5/8 TYPE C 2.  
FT SITE 85 87 CELL ATTACHMENT SITE (POTENTIAL).  
FT DISULFID 24 35 BY SIMILARITY.  
FT DISULFID 29 47 BY SIMILARITY.  
FT DISULFID 49 58 BY SIMILARITY.  
FT DISULFID 66 77 BY SIMILARITY.  
FT DISULFID 71 94 BY SIMILARITY.  
FT DISULFID 96 105 BY SIMILARITY.  
FT DISULFID 109 265  
FT DISULFID 252 256  
FT DISULFID 270 427  
FT CARBOHYD 27 27  
FT CARBOHYD 34 34  
FT CARBOHYD 59 59  
FT CARBOHYD 227 227  
FT VARSPLIC 169 221  
FT CONFLICT 19 19  
FT CONFLICT 28 28  
FT SEQUENCE 427 AA; 47411 MW; 4CBEE3A1DC4EB24 CRC64;  
  
Query Match 8.3%; Score 635; DB 1; Length 427;  
Best Local Similarity 37.9%; Pred. No. 3.8e-34;  
Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;  
  
QY 1105 ECLIGHLHAGMSTLFLVYSNK-----CQPLGMSAGHIRDFQITASQY-- 1149  
Db 76 EQQVTDSDSHRG--DVFIQVICKPLGVGHCETCTCTSPGLGQTGAIDQSASSMHLG 133  
QY 1150 ----GQWAPKLARLHYSGINAWST----KEPFSWIKVLLAPMIHIGKTQAGKQKFS 1201  
Db 134 FMGLQWAPLALRLHQTGIVNAWTSNGYDKNP--WTQVNLKRMWVTVGVVTOGASRAGA 191  
QY 1202 LVISQFIIMWLDGKKWQYRGNSTGTLVFPNGVDSGIGKNIENPPIIARVIRLHPH 1261  
Db 192 EVLKFKVAYSDGRQFQIYVAGRSGDKIFGVNNSKLNLPDTPLETQIVRLVPII 251  
QY 1262 YSIRSLRMLMGCDLNSCMPLGMESKAISDAQITASSYFTN---MFAWTSPSKARLHL 1318  
Db 252 CHRGCRLRPELLGCELNGCTEPLGLKDNTPNKQITASSYKTTWGLSAFSPFYVARLDN 311  
QY 1319 QGRSNARPOVNPKEWLDQVDFQTKMTKVTGVTQGVKSLTSMYKVEFLISSQDGHQWT 1378  
Db 312 QCKFNAWTAQTSASEWLQIDLGSKQKRVGTIITQGARDFGHIQYVAAYRVAYGDDGVTTW 371  
QY 1379 LFFQNG--KVYFQGNODSFTPWNSLDPLLTRYLRIHPQSWVHOLARMEVLGC 1432  
Db 372 EYKDFGASESKIFPGNMNDNSHKKNIFETFPQARFVRIQPVANHNRIITURVELLGC 427  
  
RESULT 14  
MFGM HUMAN  
ID MFGM HUMAN STANDARD; PRT; 387 AA.  
AC Q08431;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-EB) (HMFG)  
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].  
GN MFGEB.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast, and Breast carcinoma;  
RX MEDLINE=96213908; PubMed=8639264;  
RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;  
RT "Cloning and sequence analysis of human breast epithelial antigen  
BA46 reveals an RGD cell adhesion sequence presented on an epidermal  
growth factor-like domain.";  
RL DNA Cell Biol. 15:281-286(1996).  
RN [2]  
RP SEQUENCE OF 170-387 FROM N.A.  
RC TISSUE=Mammary Gland;  
RX MEDLINE=91371351; PubMed=1909932;  
RA Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,  
RA Ceriani R.L.;  
RT "A Mr 46,000 human milk fat globule protein that is highly expressed  
in human breast tumors contains factor VIII-like domains.";  
RL Cancer Res. 51:4994-4998(1991).  
RN [3]  
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.  
RC TISSUE=Milk;  
RX MEDLINE=98194924; PubMed=9535276;  
RA Gufrida M.G., Cavalletto M., Giunta C., Conti A.,  
RA Godovac-Zimmermann J.;  
RT "Isolation and characterization of full and truncated forms of human  
breast carcinoma protein BA46 from human milk fat globule membranes.";  
RL J. Protein Chem. 17:143-148(1998).  
RN [4]  
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.  
RX MEDLINE=99342076; PubMed=10411933;  
RA Haeggqvist B., Naeslund J., Sletten K., Westermark G.T., Mucchiano G.,  
RA Tjernberg L.O., Nordstedt C., Engstrom U., Westermark P.;  
RT "Medin: an integral fragment of aortic smooth muscle cell-produced  
lactadherin forms the most common human amyloid.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).  
RN [5]  
RP CHARACTERIZATION.  
RX MEDLINE=97405885; PubMed=9260929;  
RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;  
RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein  
expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp  
(RGD)-dependent cell adhesion.";  
RL DNA Cell Biol. 16:861-869(1997).  
CC - FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS  
SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.  
CC - FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.  
CC - SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.  
CC - TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC  
MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.  
CC - PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT  
AMINO ACID 284 AND 273.  
CC - SIMILARITY: Contains 1 EGF-like domain.  
CC - SIMILARITY: Contains 2 F5/8 type C domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC  
CC EMBL; U58516; AAC50549.1; -.  
CC EMBL; S56151; AAB19771.1; -.  
CC PIR; A47285; A47285.  
CC HSP; P08709; IBF9.  
CC Genew; HGNC:7036; MFGEB.  
CC MIM; 602281; -.  
CC GO; GO:0007155; P:cell adhesion; TAS.  
CC GO; GO:0007048; P:oncogenesis; TAS.  
CC InterPro; IPR001438; EGF_II.
```

```

DR InterPro: IPR006209; EGF like.
DR InterPro: IPR000421; F58 C.
DR InterPro: IPR006210; IEGF.
DR Pfam: PF00008; EGF; 1.
DR Pfam: PF00754; F5 F8 type C; 2.
DR PRINTS: PR00010; EGFELOOD.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00231; FAS8C; 2.
DR PROSITE: PS00022; EGF; 1.
DR PROSITE: PS01186; EGF; 2; 1.
DR PROSITE: PS01285; FAS8C; 1; 2.
DR PROSITE: PS01286; FAS8C; 2; 2.
DR PROSITE: PS00022; FAS8C; 3; 2.
DR SIGNAL: Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
KW SIGNAL
FT CHAIN 1 23
FT CHAIN 24 387
FT CHAIN 202 387
FT CHAIN 268 317
FT CHAIN 268 67
FT DOMAIN 24 67
FT DOMAIN 70 225
FT DOMAIN 230 387
FT SITE 46 48
FT SITE 27 38
FT DISULFID 32 55
FT DISULFID 57 66
FT DISULFID 70 225
FT DISULFID 212 216
FT DISULFID 230 387
FT CARBOHYD 238 238
FT CARBOHYD 325 325
FT CARBOHYD 329 329
FT CARBOHYD 350 350
FT SEQUENCE 387 AA; 43123 MW; 2EE6571DEC83782D CRC64;

Query Match 7.6%; Score 588; DB 1; Length 387;
Best Local Similarity 37.3%; Pred. No. 3.9e-11;
Matches 132; Conservative 69; Mismatches 125; Indels 28; Gaps 9;

Qy 1093 EMILPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGNASGHIRDFOITASG----- 1147
Db 48 DVFPF-----YTCTCLKG---YAGNHC-----ETKVEPLGNGNANSQIASSVRVTF 95

Qy 1148 -QYGWAPKLARLHYSGINAW--STKEPFSWIKVLLAPMIHGHTQGARQKFSLYI 1204
Db 96 LGLQHWPELARLNAGMNAWTPSSNDNPFQIQLNRRMWTGVVTCASRLASHEYL 155

Qy 1205 SQFILMYSLDGKKWOTYRGNSTGLVMYFGNVDSGGIKHNIENPPIIARVIRLHPHTYSI 1264
Db 156 KAFVAYSLNGHEFD-FIHVNKKHKEFGVGNWKNNAVHNLFTPEVAQYVRLYPTSCHT 214

Qy 1265 RSTLRMELMGCDLNSCMPLGMSKAISDAQITASSYF----TNMFATWSPSKARLHLOQ 1320
Db 215 ACTLRPELLGCELGCANPLGLKNNSPDKQITASSYKWTGLHLF-SWNPVSARLDKQG 273

Qy 1321 RSNAPQVNPKNPELQVDKTKVTGVTQGVKSLLTSMYKFEFLISSQGHQWTLF 1380
Db 274 NFNAWAGSYGNDQLQVLDLSSKEVGTITQGARFSGVQFVASYKVAYSNDNSANWTEY 333

Qy 1381 F--QNGKVKVFGQNDSTFPVNSLDPPLTLRYLRHPSQVHVHQLALMEVLG 1432
Db 334 QDPRTGSKLFPGNWHDHSHKQNLFTFPIILARYVIRLFPVAWHNRRLRLLELGC 387

RESULT 15
NRP2 HUMAN
ID NRP2 HUMAN STANDARD; PRT; 931 AA.
AC O60462; O14820; O14821;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165
receptor 2).
GN NRP2 OR VEGF165R2.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A0 AND A17).
RX MEDLINE=97470888; PubMed=9331348;
RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
RT "Neuropilin-2, a novel member of the neuropilin family, is a high
RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT III.";
RL Neuron 19:547-559(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A22).
RX TISSUE=Breast;
RA Medline=98188099; PubMed=9529250;
RT "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid
RT form of vascular endothelial growth factor (VEGF) and of placenta
RT isoform-specific receptor for vascular endothelial growth factor.";
RL Cell 92:735-745(1998).
RN [3]
RP CHARACTERIZATION
RX MEDLINE=20309748; PubMed=10748121;
RA Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
RT "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid
RT form of vascular endothelial growth factor (VEGF) and of placenta
RT growth factor-2, but only neuropilin-2 functions as a receptor for
RT the 145-amino acid form of VEGF";
RL J. Biol. Chem. 275:18040-18045(2000).
CC -1- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
CC AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
CC -1- SUBUNIT: NEUROFILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
CC NEUROFILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN 3C RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=A22;
CC IsoId=O60462-1; Sequence=Displayed;
CC Name=A0;
CC IsoId=O60462-2; Sequence=VSP_004342;
CC Name=A17;
CC IsoId=O60462-3; Sequence=VSP_004341;
CC -1- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
CC -1- SIMILARITY: Contains 2 CUB domains.
CC -1- SIMILARITY: Contains 2 F5/8 type C domains.
CC -1- SIMILARITY: Contains 1 MAM domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; AF02859; AAC51788.1; -
DR ENBL; AF02860; AAC51789.1; -
DR ENBL; AF016098; AAC12922.1; -
DR HSP; P12259; ICZT.
DR Genew; HGNC:8005; NRP2.
DR MIM; 602070; -
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0004872; F:receptor activity; TAS.
DR GO; GO:0005021; F:vascular endothelial growth factor receptor. . .; TAS.
DR GO; GO:0007411; P:axon guidance; TAS.
DR InterPro; IPR000859; CUB domain.
DR InterPro; IPR000421; FAS8 C.
DR InterPro; IPR000998; MAM domain.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR Pfam; PF00629; MAM; 1.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 2.

```


DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C.1; 2.
DR PROSITE; PS01286; FA58C.2; 2.
DR PROSITE; PS01287; FA58C.3; 2.
DR PROSITE; PS00222; FA58C.3; 2.
DR PROSITE; PS00222; FA58C.3; 2.
KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW Alternative splicing.
FT SIGNAL 1 20 OR 22 (POTENTIAL).
FT CHAIN 21 931 NEUROFILIN-2.
FT DOMAIN 21 864 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 865 889 POTENTIAL.
FT DOMAIN 890 931 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 28 142 CUB 1.
FT DOMAIN 149 267 CUB 2.
FT DOMAIN 277 427 FS/8 TYPE C 1.
FT DOMAIN 434 592 FS/8 TYPE C 2.
FT DOMAIN 642 802 MAM.
FT DOMAIN 671 674 POLY-SER.
FT DISULFID 28 55 BY SIMILARITY.
FT DISULFID 83 105 BY SIMILARITY.
FT DISULFID 149 175 BY SIMILARITY.
FT DISULFID 208 230 BY SIMILARITY.
FT DISULFID 277 427 BY SIMILARITY.
FT DISULFID 434 592 BY SIMILARITY.
FT CARBOHYD 152 152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 157 157 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 839 839 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 809 813 Missing (in isoform A17).
FT VARSPLIC 809 830 Missing (in isoform A0).
FT CONFLICT 602 602 /FTID=VSP 004342.
FT SEQUENCE 931 AA; 104830 MW; 270CBAB69A0A797C CRC64;
Query Match 6.1%; Score 469.5; DB 1; Length 931;
Best Local Similarity 28.5%; Pred No. 8.1e-23;
Matches 166; Conservative 87; Mismatches 185; Indels 145; Gaps 32;
QY 934 DEFDC-AWAYPSVDLEK-----DVHSLGILGLLVCHNTL-----NPAHGRQVTVQBEFA 983
DB 79 EKHDCKYDFEIRDDGSDSADLLKGHCNIAPTTISSGSMYIKFTSDYARQGA--GFS 136
QY 984 LFFETIDFKSWYFTENMERNCRAPNIQMEDPTKENYRPHAINGYIMDTLPLGLVMAQD 1043
DB 137 LRYEIF-KTGS-----EDCSKFTSP-NGTIESPGFPEKYP----- 170
QY 1044 QRIRWYLLSMGNSNIHSHFSGHVFTVRKKEEYKALYNLYPGVFETVEMLPKAG-- 1100
DB 171 -----HNLDCFTILAKPKMEIILQFL---IPD-LEHDPLOVGECD 207
QY 1101 -----IW-----RVECLIGH-----LHAGMT-----LFLV 1122
DB 208 CKYDMLDIWDGIPHYGLIGKYGKTPSELRSSTGILSLTPTDMAVAKDGFSAARYLV 267
QY 1123 YSN-----KCQTPLGMASGHIRDFOITAGQY--QWAPKLARLHYSGINAW-----ST 1170
DB 268 HOEPLLENQCNVPLGMSGRIANEQISASTYSVSGRWTPQOSRLH--GDDNGWTFNLDN 325
QY 1171 KEPFSWIKVDLLAPMIIHGKTKQGA--ROKFSYLSIOFIIMYSLDGKKWQTYRGNSTGT 1228
DB 326 KE---YLQVDLRFLLTALATQGAISRETQNGYVTKYLEVSTNGEDMWYRHGKNH- 381
QY 1229 LMVFFGNVDSGKIHNIFFPIIARYIRLHPHYISIRTLRMLMGCDLNS--CSMPLGM 1286
DB 382 -KVFQANNDATVNLKHLAPLITFVRIRPQVTHSGIALRLLEFGCRVTDAPCSNMLGM 440
QY 1287 ESKAISDAQITASSYFTNFWATPSKARLHQGSNAW---RPQVNNPKWLQVDFOKT 1343
DB 441 LSGLIADSQISASS--TOEY-LWSPSAARL-VSSRS-GWFFRIPQAPQGEWLQVDLQTP 495

Search completed: December 9, 2003, 16:54:09
Job time : 62 secs

QY 1344 MKVTGVTTQG-----VKSLTSMYVKEFLISSODGHQWTLFFONGKV---KVFQGNOD 1394
DB 496 KTVKGVIIQARGGDSITAVEARAFVRKFKVSYSLNGKDWZ--YIQQPRTQPKLPEGNMH 554
QY 1395 SFTPVVNSLDPPPLLTRYLRHQPQSVVHQ--IALRMEVLGCEAQD 1436
DB 555 YDTPDIRRFD-PIPAQYVRYVPERWSPAGIGRWLEVLGCDWTD 596

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 16:50:01 ; Search time 54 Seconds
(without alignments)
6871.845 Million cell upd

Title: US-10-006-091-1

Perfect score: 7691
Sequence: 1 ATRRYVLGAVELSWDYMOSD.....WVHQTALRMVVLGCEAODLY 1438

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000
Maximum DB seq length: 200000000[illegible]

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23:*

```

1:  sp_archaea:*
2:  sp_bacteria:*
3:  sp_fungi:*
4:  sp_human:*
5:  sp_invertebrate:*
6:  sp_mammal:*
7:  sp_mmc:*
8:  sp_organelle:*
9:  sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score		Query %		DB	ID	Description
	Match	Length	Match	Length			
1	6272	81.5	2343	6	O18806	O18806 canis famil	
2	6231	81.0	2343	6	O62730	O62730 canis famil	
3	2486.5	32.3	2119	13	Q90X47	Q90X47 brachydactyl	
4	2386.5	31.0	2224	4	O43737	O43737 homo sapien	
5	2369	30.8	2183	11	O88783	O88783 mus musculus	
6	1799	23.4	1157	11	O92024	O92024 mus musculus	
7	1784	23.2	1157	11	O920H8	O920H8 rattus norv	
8	1747	22.7	1158	4	Q9BQS7	Q9BQS7 homo sapien	
9	1742	22.6	1158	4	O9C058	O9C058 homo sapien	
10	1710	22.2	1104	4	O75180	O75180 homo sapien	
11	1709.5	22.2	1084	11	Q9X197	Q9X197 rattus norv	
12	1618	21.0	1048	6	Q9XT27	Q9XT27 ovis aries	
13	1534.5	20.0	355	11	O8BQ43	O8BQ43 mus musculus	
14	1285	16.4	847	11	O8C4S2	O8C4S2 mus musculus	
15	1221	15.9	782	4	O75659	O75659 homo sapien	
16	1104	14.4	216	4	Q14286	Q14286 homo sapien	

17	1046.5	13.6	626	13	Q90ZT2	Q90zt2 brachydanio
18	911	11.8	407	13	Q8AYE0	Q8ave0 brachydanio
19	827	10.8	503	11	Q8BV37	Q8bv37 mus musculus
20	779.5	10.1	1142	10	Q8LL91	Q8ll91 chlamydomon
21	664	8.6	453	11	Q9RlX9	Q9rlx9 mus musculus
22	662.5	8.6	480	4	Q8NG10	Q8ng10 homo sapien
23	662.5	8.6	480	4	O43854	O43854 homo sapien
24	661	8.6	426	11	Q9WTS3	Q9wts3 mus musculus
25	655.5	8.5	470	11	Q8C4U8	Q8c4u8 mus musculus
26	655.5	8.5	480	11	O35474	O35474 mus musculus
27	655.5	8.5	480	11	Q8CBF7	Q8cbf7 mus musculus
28	514.5	6.7	363	6	Q77718	Q77718 equus caball
29	488	6.3	335	4	Q9BTL9	Q9btl9 homo sapien
30	459.5	6.1	901	4	Q9H2E4	Q9h2e4 homo sapien
31	469.5	6.1	901	4	Q9H2D5	Q9h2d5 homo sapien
32	469.5	6.1	906	4	Q9H2D4	Q9h2d4 homo sapien
33	469.5	6.1	906	4	Q9H2E3	Q9h2e3 homo sapien
34	466.5	6.1	384	11	Q8C8K0	Q8c8k0 mus musculus
35	482.5	6.0	926	11	Q80ZY7	Q8qzy7 mus musculus
36	450.5	5.9	921	11	Q9QX38	Q9qx38 rattus norv
37	446.5	5.8	919	13	Q8UVR0	Q8uvr0 gallus gall
38	446.5	5.8	936	13	Q8UVR0	Q8uvr0 gallus gall
39	429.5	5.6	609	4	Q96190	Q96190 homo sapien
40	429.5	5.6	644	4	Q961H5	Q961h5 homo sapien
41	429.5	5.6	704	4	Q9H2E1	Q9h2e1 homo sapien
42	415.5	5.4	923	13	Q8QFX6	Q8qfx6 brachydanio
43	412.5	5.4	923	13	Q8AXP1	Q8axp1 brachydanio
44	411.5	5.4	838	5	O76470	O76470 lytechinus
45	383.5	5.0	555	4	Q9H2E2	Q9h2e2 homo sapien

ALIGNMENTS

RESULT 1

018806	PRELIMINARY;	PRT;	2343 AA.
AC	018806;		
AC	018806;		
DT	01-JAN-1998 (T-EMBLrel. 05, Created)		
DT	01-JAN-1998 (T-EMBLrel. 05, Last sequence update)		
DT	01-MAR-2003 (T-EMBLrel. 23, Last annotation update)		
DE	Factor VIII.		
DE	F8.		
GN	F8.		
OS	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutele-		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis		
OC	NCBI_TaxID=9615;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Liver;		
RA	Cameron C., Noley C., Hoyle S., McGlynn L., Hough C., Kan-		
RA	Giles A., Lillcrap D.		
RT	"The canine factor VIII cDNA and 5' flanking sequence."		
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBD databases.		
CC	-1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.		
CC	EMBL; AF016234; AB87412.1; -.		
DR	HSP; P00451; 1CFG.		
DR	InterPro; IPR001117; Cu-oxidase.		
DR	InterPro; IPR000421; PA58 C.		
DR	Pfam; PF00394; Cu-oxidase; 3.		
DR	Pfam; PF00754; F5_F8 type C; 2.		
DR	SMART; SM00231; FA58C_2.		
DR	PROSITE; PS01285; FA58C_1; 2.		
DR	PROSITE; PS01286; FA58C_2; 2.		
DR	PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.		
DR	SEQUENCE 2343 AA; 265829 MW; A854FAE571C3B399 CRC64;		
SQ			

Query Match 81.5%; Score 6272; DB 6; Length 2343;
Best Local Similarity 53.5%; Pred. No. 0;
Matches 1245; Conservative 82; Mismatches 101; Indels 900; Gaps 4;

QY 1 ATRYYLGAVELSWDYMOSD-LGELPDARFPRVPKSPFNSTVVKTLFVEFTVHLP 59
 +
 A T R Y Y L G A V E L S W D Y M O S D - L G E L P D A R F P R V P K S P F N

Db	20	ATRKYLGAVELSWDYMOSDLSALHADTSFSSRVPGLPLTTISVTYKTVFVFETDDLF	79
Qy	60	NIAPRPPWGLLPTTQAEVYDTWVITLKNMASHVSLHAGVSYWKASEGAEYDDQTS	119
Db	80	NIAPRPPWGLLPTTQAEVYDTWVITLKNMASHVSLHAGVSYWKASEGAEYDDQTS	139
Qy	120	QREKEDDKVFPFGSHTVVOVLKENGPMASDPLCLTVSYLSHVDLVKDLNSGLIGALLVC	179
Db	140	QKKEEDNVIPGSHTYVVOVLKENGPMASDPLCLTVSYLSHVDLVKDLNSGLIGALLVC	199
Qy	180	REGSLAKEKTQTLHKFILLFAVDEGKSHSETKNSLMQDRDAASARAWPKMHTVANGYN	239
Db	200	KEGSLAKERTOTLQEVLLFAVDEGKSHSETNASLTQ-----AEAQHELTINGYN	253
Qy	240	RSLEGLIGCHRSYVHVIGMGTTPVHSHIFLGHTFLVRNHRQASLEISPTIFLTAQTL	299
Db	254	RSLEGLTVCHRSYVHVIGMGTTPVHSHIFLGHTFLVRNHRQASLEISPTIFLTAQTF	313
Qy	300	LMDLGQFLFCHISSHOHQGMBAVVKVDSCEPPEPQLRMKNREAEYDDDLTDSMDVVR	359
Db	314	LMDLGQFLFCHIPSHOHQGMBAVVKVDSCEPPEPQLRMKNED-KOTDDGLYDSMDVVS	372
Qy	360	FDDNSPFSFIQIRSAVAKHPKTVVHYIAAEEDWDYAPLVAPDDRSYKSYQLNNGPQRI	419
Db	373	FDDSSSPFIQIRSAVAKHPKTVVHYIAAEEDWDYAPSGPTNDRSHKNLYLNGPQRI	432
Qy	420	GRKYKVRPMAYTDEFKTRATQESGILGPLLYGEVGDTLIIIFNQAASRPYNTYPHG	479
Db	433	GRKYKVRPMAYTDEFKTRATQESGILGPLLYGEVGDTLIIIFNQAASRPYNTYPHG	492
Qy	480	ITDVRPLYSRRLPKGVKHLKDPILPGEIFPKYKWTVTVEDGPKSDPRCLTRYSSFVNM	539
Db	493	INVTFLHTGRLLPKGVKHLKDPILPGEIFPKYKWTVTVEDGPKSDPRCLTRYSSFINL	552
Qy	540	ERDLASGLIGPLLI CYKESVDQRGNOIMSDKRNVLFSVFDENRSWYLTENIORFLPNEA	599
Db	553	ERDLASGLIGPLLI CYKESVDQRGNOIMSDKRNVLFSVFDENRSWYLTENMQRFLPNAD	612
Qy	600	GVULEPPEQASIMHSINGYVDSLSQLSVCLHEVAYWYILSGAOTDFLSVFSSGYTFK	659
Db	613	VQPHDPPEQLSIMHSINGYVDPNLQLSVCLHEVAYWYILSGAOTDFLSVFSSGYTFK	672
Qy	660	HKVYEDTTLTPFPSETVFMSENPLATLGCHNSDFNRGNTALLKYSSCDKNTGDYX	719
Db	673	HKVYEDTTLTPFPSETVFMSENPLWVLGCHNSDFNRGNTALLKYSSCNRNDDYY	732
Qy	720	EDSYEDISAYLLSKNNAIPRSP-----	742
Db	733	EDTYEDIPTLLNENNVIKPRFSQNSRHPSTKEKOLKATTPPENDIEKIDLOSGERTOL	792
Qy	743	-----	742
Db	793	IKAQSVSSDLLMLLQONTPRGLFSLDREATDRADHSGAIGERNKGPPEVASIRPEL	852
Qy	743	-----	742
Db	853	RHSEDEFTPEBELQLRINENGTNTVELKKLDLKISSSDSLMSTSPITPSDKLAAATE	912
Qy	743	-----	742
Db	913	KTGSLGPPNNSVHFNSHLGTIVFGNNSSHLIQSGVPLELSEEDNSKLLLEAPLMNIQESS	972
Qy	743	-----	742
Db	973	LRENVLMSNRILFKEERIRGPASLIKDNALFKVNISSVKTNRAPVNLTTNRKTRVAIPT	1032
Qy	743	-----	742
Db	1033	LIENSTSVQDMLERNTEFEKVTSLIHNETFMDRNTTALGLNHSNKTTLISKVEMAH	1092
Qy	743	-----	742
Db	1093	QKEDVPVLAENPDLSSSKIPFLPDWIKTHGKNSLSSEQRSPKQTLTSLGSEKSKVQON	1152
Qy	743	-----	742
Db	1153	FLSEKVVVGEDEFTKOTELQEBFPNNKSIFFANLANVQENDTYNQEKSPBEIERKXL	1212
Qy	743	-----	742
Db	1213	TQENVALPOAHWTIGTKNFKNLFLLLSTKONVAGLEEQPYTPILOTRSLNDSPHSEGIH	1272
Qy	743	-----	742
Db	1273	MANFSKIREANLEGLGNQTNQVBERFPSTTRSSNASQHVITQRKRSILKQPRLSQGR	1332
Qy	743	-----	742
Db	1333	KPERKVIANDTSTQMSKNMNYLAQGLTQIETNEYKEKRAITQSPSLSDCSMRNHVTQMND	1392
Qy	743	-----	742
Db	1393	SALPVAKESASPSVRHTDLTKIPSHNSHLPASACNYYTFRBRTSGVBQSHFLQEAEN	1452
Qy	743	-----	742
Db	1453	NLSLAFVTLGITEGQKFSGLCKSATNQPMYKXLENVTLLOPGLSETSKVELLSQVHVD	1512
Qy	743	-----	742
Db	1513	QEDSPPTKTSNDSPGHLDLMGKIFLQKQGVKMNKNSPGKVPFLKWATESSEKIPSKL	1572
Qy	743	-----	742
Db	1573	LGVLAWNHYDQIIPSEHWKSKSQTNTAFKRKDTIPLGPCENNDSAAINEGQDKPQ	1632
Qy	743	-----SONPPVLKRHQREITRTTLOSDBEIDYDDTISVEMKEDFDIYD	787
Db	1633	REAWAKQEGEPGLCSQNPVSKHQREITVTTLQPEEDKFYDDTTFSEMKREBFDIYG	1692
Qy	788	EDENSPRSFQKTRHYFIAAVERLDYGMSSPHVLNRNAQSGSVQPKVVFQFRTDG	847
Db	1693	DYENQGLRSFQKTRHYFIAAVERLDYGMSSPHVLNRNAQSGSVQPKVVFQFRTDG	1752
Qy	848	SFTQPLRGELNEHLGLIGPYIRAEVEDNIMVTFNQAASRPYSFSSLSIYSEEDORQAE	907
Db	1753	SFTQPLRGELNEHLGLIGPYIRAEVEDNIMVTFNQAASRPYSFSSLSIYSEEDORQAE	1812
Qy	908	PRKNFVKNETKTYFWKQVHMAPTKDFCDKAWAYFSDVDLEKDVHSLGLIGLIVCHTN	967
Db	1813	PRKFNVENETKTYFWKQVHMAPTKDFCDKAWAYFSDVDLEKDVHSLGLIGLIVCHTN	1872
Qy	968	TLPNAGRQVTVQEBFALFTTIPDETYSWYFTENMERNCRAPCNIQMEDPTPKENYRFAH	1027
Db	1873	TLPNAGRQVTVQEBFALFTTIPDETYSWYFTENMERNCRAPCNIQMEDPTPKENYRFAH	1932
Qy	1028	NGYIMDTLPGINMAODQIRWYLLSMGNSNETHSFSGHVTWPKKBYKXALVNLVPG	1087
Db	1933	NGYIMDTLPGINMAODQIRWYLLSMGNSNETHSFSGHVTWPKKBYKXALVNLVPG	1992
Qy	1088	VFETVEMLPFSKAGIWRVCECLIGEHLHAGMSTLFLVYSNKCOTPLGMAHGHRDFOITAG	1147
Db	1993	VFETVEMLPFSKAGIWRVCECLIGEHLHAGMSTLFLVYSNKCOTPLGMAHGHRDFOITAG	2052
Qy	1148	QYGQWAPKARLHYSGSINAWSTKEPFSWIKVLDLAPMIHIGITQAGARQKFSLSYISQF	1207
Db	2053	QYGQWAPKARLHYSGSINAWSTKEPFSWIKVLDLAPMIHIGITQAGARQKFSLSYISQF	2112
Qy	1208	LIYSLDCKKQVTVRGSTGTLMVFFGNVDSGKKNIPNPIIARYIRLPHTHYSIRST	1267
Db	2113	LIYSLDCKKQVTVRGSTGTLMVFFGNVDSGKKNIPNPIIARYIRLPHTHYSIRST	2172
Qy	1268	LRMELMGCDLSCSNMPLGMSKALSDAQITASSYFTNMFAWSPSKARLHLOGRNWRP	1327
Db	2173	LRMELMGCDLSCSNMPLGMSKALSDAQITASSYFTNMFAWSPSKARLHLOGRNWRP	2232

```
QY 1328 QVNNPEVLOYDFQKTMKVTTGTTQGVKSLLTSMTVKSEFLISSQDGHQWTLFPONGKVK 1387
Db 2233 QANNPEVLOYDFRKTMTVTGTTTQGVKSLLTSMTVKSEFLISSQDGHQWTLFPONGKVK 2292

QY 1388 VFQGNQDSFTPVNSLDPLLRXYRIHPQSVVHQAIALRMEVLGCEAQ 1435
Db 2293 VFQGNRDSSTPVNRLEPPLVARYVRLHPQSVVHQAIALRMEVLGCDTQ 2340

RESULT 2
O62730 PRELIMINARY; PRT; 2343 AA.
AC O62730;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Factor VIII.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN SEQUENCE FROM N.A.
RP TISSUE-Kidney, and Spleen;
RA Gordy P.W., Bowen R.A.;
RT "Characterization of the canine factor VIII cDNA.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; AF049489; AAC05384.1; --
DR HSSP; P00451; 1CFG
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8 type C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.
SQ SEQUENCE 2343 AA; 265613 MW; F612D744ADAADD99 CRC64;

Query Match 81.0%; Score 6231; DB 6; Length 2343;
Best Local Similarity 53.2%; Pred. No. 0;
Matches 1238; Conservative 85; Mismatches 105; Indels 900; Gaps 4;

QY 1 ATRRYLGAVELSDYMQSD-LGELPVDAREPPVPKSPFENTSVYKTLFVFTVHLF 59
Db 20 ATRKYLGAVELSDYMQSDLLSALHADTSFSSRVGSLPTTSVYKTVFVFTDGLF 79

QY 60 NIAKPRPPMGLLPTTQAEVYDVTVIILKNASHPVSLHAGVSYWKASGAEYDDQTS 119
Db 80 NIAKPRPPMGLLPTTQAEVYDVTVIILKNASHPVSLHAGVSYWKASGAEYDDQTS 139

QY 120 QREKDDKVPFGSHYVYVQVLEKNGPMASPLCTYSYLSHVDLVKDLNSGLIGALLVC 179
Db 140 QREKDDNVIPGSHYVYVQVLEKNGPMASPLCTYSYLSHVDLVKDLNSGLIGALLVC 199

QY 180 REGSLAKEKTLTKHFTILLFAVFDGKSWHSETKNLSMQDRDAASARAPKQHTVNGYVN 239
Db 200 KEGSLAKERTQLQEFVLLFAVFDGKSWHSETNASITQ-----AEQHELHTINGYVN 253

QY 240 RSLPGLGCHRSKYVHYVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISITFLTAQTL 299
Db 254 RSLPGLATVCHRSKYVHYVIGMTTPEVHSIFLEGHTFLVGNHRQASLEISITFLTAQTF 313

QY 300 LMDLQGFLLFCHISSHQDGHAEYKVDSCPEEPQLRMKNNEAEYDDDLTDSMDVVR 359
Db 314 LMDLQGFLLFCHIPSHQDGHAEYKVDSCPEEPQLRMKNNEAEYDDDLTDSMDVVS 372

QY 360 FDDNNSPSFIQIRSVAKKHPKTVHYIAAEEDWDYAPLVLPDDRYSYKQYLNNGPQRI 419
Db 373 FDDSSSPPFIQIRSVAKKHPKTVHYIAAEEDWDYAPSGPTPNDRSHKNLYLNNGPQRI 432

QY 420 GRKYKVRFWAYTDTETFKTRAIQHESGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHG 479
```

```
Db 433 GKTKYKVRFWAYTDTETFKTRAIQHESGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHG 492
QY 480 ITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKKTVTTVEDGPTKSDPRCLTRYYSFVNM 539
Db 493 INYVPLHTGLRPLKGVKHLKDFPILPGEIFKYKKTVTTVEDGPTKSDPRCLTRYYSFNL 552
QY 540 ERDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRFPLNPA 599
Db 553 ERDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRFPLNPA 612
QY 600 GVQLEDEPFOASNMHSINGVYVDSLOLSVCLHVAWYIILSIGAOTDPLSVFSGYTFK 659
Db 613 VVQPHDPEFQLSNMHSINGVYVDFNLQLSVCLHVAWYIILSVGAOTDPLSVFSGYTFK 672
QY 660 HKMYVEDTLTLFPFSGTVMFMSNPGIWLIGCHNSDFRNRGMTALLKYSSCDKMTGY 719
Db 673 HKMYVEDTLTLFPFSGTVMFMSNPGIWLIGCHNSDFRNRGMTALLKYSSCDKMTGY 732
QY 720 EDSYEDISAYLLSKNNAIEPRSF----- 742
Db 733 EDYTEDIPTPLANNVVKPRSFQNSRHPSTKEKQLKATTPPENDIEKIDLQSGERTQL 792
QY 743 ----- 742
Db 793 IKAQSVSSDILLMLLQGNPTPRGLFGLDREATDRADHSGAERNKGPPEVASLRPEL 852
QY 743 ----- 742
Db 853 RHSEDRFTPEPELQRLNENLGNVTNTVELKKDLKISSSDSLMTSPTIPSDKLAATE 912
QY 743 ----- 742
Db 913 KTGSLGPPNMSVHFNGHLGTVFCGNSSHLIQSGVPLELSEEDNDSKLEAPIANIQESS 972
QY 743 ----- 742
Db 973 LRENVLSMESNRLFKEERIRGPASLIKONALFKVNISSVKNRAPVNLITTRKTRVAIPT 1032
QY 743 ----- 742
Db 1033 LLIENTSVQMDIMLERNTEFKEVTSLIHNTEFMDRNTTALGLNHNKTLTSKNVMAH 1092
QY 743 ----- 742
Db 1093 QKKEDPVPLAENPDLSSKIPFLPDWKTHGKNLSSEORPPSPKQLTSLGSEKSVKQDN 1152
QY 743 ----- 742
Db 1153 FLSEKVVVGEDEFTKDELQEI FPNKNSIFFANLANVQENDTYNQEKKSLEERKEKL 1212
QY 743 ----- 742
Db 1213 TQENVALPOAHTWIGTKNFKNLFLSLTKQNVAGLEEQYTPILQDTRSLNDSHSGI 1272
QY 743 ----- 742
Db 1273 MANFSKIREANLEGLNQTNQVVERPPSTRSSNASQHVITQGRKRSILKQPLSQGEI 1332
QY 743 ----- 742
Db 1333 KPERKVIANDTSTQWSKNMNYLAQGTLTQIYEYNEKEKRAITQSPSLDCSMRNHVITQ 1392
QY 743 ----- 742
Db 1393 SALPVAKESASPSVRHTDLTKIPSOHNSHLPASACNYTPRERTSGVQEGSHFLQAKRN 1452
QY 743 ----- 742
Db 1453 NLSLAFVTLGITGQOKFSSLGKATNQPMYKLENTVLLQPLGSETSKVELLSQVHVD 1512
QY 743 ----- 742
```

Db 1513 QEDSPPTKTSNDSPGHLDMGKI FLQKTQGPVKMNKTSNPGKVPFLKWATESSEKIPSKL 1572
QY 743 ----- 742
Db 1573 LGVLAWDNHYDTQIPSEEWKSKQSQTNTAFKRKDTILPLGPCENNDSATAINEGQDRKP 1632
QY 743 -----SQNPVLKRHOREITRITTLQSDQBEIDYDDTISVEMKEDFDYD 787
Db 1633 REAMWAKQEPGRCLCSQNPVSKHOREITVTTLQPEEDKFYDDTFSIEMKREDFDIYG 1692
QY 788 EDENSGPSRFSQKTRHYFIAAVERLWYGMSSPHVLNRNRAQSGVPQKVVQFQFTDG 847
Db 1693 DYEDQGLRSFQKTRHYFIAAVERLWYGMSSPHVLNRNRAQSGVPQKVVQFQFTDG 1752
QY 848 SFTQPLRGELNEHLGLGPGYRAEVEDNIMWTFNQAASRPSPYSSLSIYSEEDORQAE 907
Db 1753 SFTQPLRGELNEHLGLGPGYRAEVEDNIMWTFNQAASRPSPYSSLSIYSEEDORQAE 1812
QY 908 PRKNVFNKNETKYTFWKVQHMAPTKDEPDKAMAYFSDVDLEKDVHSGLIGPLLVCHTN 967
Db 1813 PRKFWFNNETKLYFWKVQHMAPTKDEPDKAMAYFSDVDLEKDVHSGLIGPLLVCHTN 1872
QY 968 TLNPAHQGVTVQEPALFETIPDETQSWYFTENMEERNCRAPCNIQMEDPTFKENYRFAI 1027
Db 1873 TLNPAHQGVTVQEPALFETIPDETQSWYFTENMEERNCRAPCNIQMEDPTFKENYRFAI 1932
QY 1028 NGYINDTLPGLVMAODQIRWYLLSGNSNENIHSIFSGHVTFRKBEYKXALNLYPG 1087
Db 1933 NGYVXDTPGLVMAODQIRWYLLSGNSNENIHSIFSGHVTFRKBEYKXALNLYPG 1992
QY 1088 VPETVEMLPSPKAGIWRVECLIGEHLHAGNSTLFLVYSNKCOTPLGWSGHIRDFOITASG 1147
Db 1993 VPETVEMLPSPQVGIWIECLIGEHLHAGNSTLFLVYSNKCOTPLGWSGHIRDFOITASG 2052
QY 1148 QYGWAPKARLHYSGSINAMSTKPFQWIKVDLLAPMIHIGIKTQGARQKPFSSLYISQF 1207
Db 2053 QYGWAPKARLHYSGSINAMSTKPFQWIKVDLLAPMIHIGIKTQGARQKPFSSLYISQF 2112
QY 1208 IIMYSLDGKKWQTYEGNSTGLTWFFGNVDSGSIKHNIFNPDIARYIRLHPHTYSIRST 1267
Db 2113 IIMYSLDGKKWQTYEGNSTGLTWFFGNVDSGSIKHNIFNPDIARYIRLHPHTYSIRST 2172
QY 1268 LEMELMGCDLNSCMPLGMSKAISSAQITASSYFTNNFATWSPSKARLHLQGRNANRP 1327
Db 2173 LEMELMGCDLNSCMPLGMSKAISSAQITASSYFTNNFATWSPSKARLHLQGRNANRP 2232
QY 1328 QVNNPKWLQVDFQKTMKVTVGTQGVKSLLTSMYVKBFLISSQDGHOWTLFFQNGKVK 1387
Db 2233 QANNPKEWLQVDFRKTMTKVTGITTQGVKSLLTSMYVKBFLISSQDGHOWTLFFQNGKVK 2292
QY 1388 VFQGNQDSFTPVNSLDPILLTRYLRIRHQPQSVHQAIALRMEVLGCEAQ 1435
Db 2293 VFQGNQDSFTPVNSLDPILLTRYLRIRHQPQSVHQAIALRMEVLGCEAQ 2340

RESULT 3

Q90X47 PRELIMINARY; PRT; 2119 AA.
AC Q90X47;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE SC:hz2015.3 (Novel protein similar to vertebrate coagulation factor V and VIII).
GN SC:hz2015.3.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_taxid=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Lloyd D.;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBAJ databases.
CC -I- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; AL590146; CAC94896.1; --
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 2.
DR SMART; SM00754; F5_F8_type_C; 2.
DR SMART; SM00231; F56G; 2.
DR PROSITE; PS01285; FAS8C_1; 2.
DR PROSITE; PS01286; FAS8C_2; 1.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
SQ SEQUENCE 2119 AA; 240643 MW; DC0806FFA8761E6 CRC64;

Query Match 32.3%; Score 2486.5; DB 13; Length 2119;
Best Local Similarity 27.7%; Pred. No. 6.3e-177;
Matches 600; Conservative 262; Mismatches 491; Indels 815; Gaps 36;

QY 3 RRYTLGAVELSDWYMQSDLGELPVDARPPRPVPSFPNTSVVYKTLTFVFTVHLEFVIA 62
Db 28 RHYIAAVNINWDYTSQQ-----RTGQSYKVVYREYN-EGFKQP 67
QY 63 KPRPFWMLLGLPTTQAEVYDVTWITLKNMASHVSLHAGVSYWKASEGAYDDOTSORE 122
Db 68 KAHELSSGLLGLPTLRGQEGDTIIVFRNADHPGSLPHGLAYGKQSGSLYFDNTSLLE 127
QY 123 KEDDKVPPGGSHYVWQVLEKNGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCREG 182
Db 128 KNDVIOQGEHHTYQMDVTSVDTPTAADPPCITYSLSHFDIVRDYNTGLTGMPLICKKG 187
QY 183 SLAKEKTQTLH---KFILLFAVDFEGKSWHSETKNSLMQDRDAASARAPKMHVWGVN 239
Db 188 TLDDSGNQ-IHFQESVLLFGVFDENKSWY-----TGDSPPQLVNKYTINGYN 236
QY 240 RSLPGLTCHRSKSYVHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISITFLTAQTL 299
Db 237 GSPDLDICAHSKYKSWHLLGMSSELPFSVHFNGQVLIHDGHTKSAVGIIISGTATASMT 296
QY 300 LMDLQGLLFCFHISSHQHDGMEAYVVKVDSCE--EPQLRMKNNEAEYDDDLTSDMDV 357
Db 297 GVHGRMLVSSHISKHLEAGLYNIRKDCDEYAPKRLRTIEQ----- 340
QY 358 VRFDNDSPTSIQIRSAKHKPTWVHYIAAEEDWDYAPLVAPDPRSRYKSOVLNNGQP 417
Db 341 -----KKESQWYMAAEVWDYAPNPNMMDGDFSKVLKQGPQ 382
QY 418 RIGKYKVRPMAYTDEFTTR---BAIQHESGILGPLLYGEVGTLLIIFPNQASRPYN 474
Db 383 RIGKYKAVPTQYKDGKFKERABDKQRELGLGPVIRAIIRDIKIIVFNKASRPS 442
QY 475 IYPHGIITVRLYSRRLPKGVKHLKDPILPGLPEIFKYKMTVYVDEGPKSPCLTRYYS 534
Db 443 IYPHGLTIDKAAEGASYPQGN--QTSYVQGEITYTWTSTEDVPTSDPRLTRMYH 500
QY 535 SFVNMERDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSHYLTENIORF 594
Db 501 SAVDAPRDIASGLVGPLLI CKSQSLNKNVQKXADKQCHAMFTVDENKSHYQENINTY 560
QY 595 LPNPAGVQLEDPEFQASNMHMSINGYVFDLSIQ-LSVCLHVEYAVYIISIGAQTFLSVFF 653
Db 561 CSDEPKVKKDDPEFYKSNVMEHTINGYVYESQBLGFGCHGEIVTWHVSVGQDYIQTATP 620
QY 654 SGYTFKHMYVEDTLTPFPFSGETVFMENPGLMILGCHNSDFRNGMTALLKVSCKDK 713
Db 621 YGHTFELKREEDILSLFPMTGETITVMNVNIGIWLASLNSHSDTKGMRYKFKDLCEFR 680
QY 714 NTGQY-----YE-----DSYEDISAYLL----- 731
Db 681 ---DYVLEYDEGKFTAWKPPPTINEIKKEBPVAPDVVDVEYSDLPASTLNLTFFNVK 737
QY 732 -----SK 733
Db 738 DEVEIIDLTFLDQDDGLLPVVEKSLGSSNENLHNATLQSFIEHTHGLMBEGDLKDGESS 797

```
QY 734 NNAI-----EPRFS 743
Db 798 NKVLNDSTKALLETTFDNRVVALNNETDSIILDFPIVERKVSAPKPMPEPESVT 857
QY 744 QN-----PVLKRHQREIIRT-----TLQSD-- 764
Db 858 MNFKTEHINSLSERINAIPIETNTINWIEHTDFTSIPTFDGSGEMNFTLEDDTAL 917
QY 765 -----QEEI-----DYD-----DTISVEMK- 779
Db 918 LNSSESEPLQNSNENRIAFQELNNAKGDVDSDNNSVKQIKPKYNNVPSGDTLNSSKI 977
QY 780 --KEDFDIYD----- 787
Db 978 QVEEDFVLLDSSYFSEMSTMEYDVSQDVTYKGSKETAQSQELSSTKTKYSGEILLES 1037
QY 788 ----- 787
Db 1038 PDITSAPNLSSVLNRNLSNBSNETLPMSSNATFSDSTNATSDSSTATPADFSNT 1097
QY 788 ----- 787
Db 1098 TFSNATPFDGFRISQMSDSSNATLSDSSNATLSDSSNATLSDSSNATLSDS 1157
QY 788 ----- 787
Db 1158 SNATPDSNKTFSNATLFGVSYSSNTTSLDPSLESEMTEYLSSANDTIKSHSEV 1217
QY 788 -----EDENQ----- 792
Db 1218 SNTQLSSSESTENISLLYGLNASSMKNDSESEEVVILNKNHSEAILTSHLDQKE 1277
QY 793 ----- 792
Db 1278 EHWGESKHELVHKLPHENKMYKDKSAANSKPKLEKKKYQVRKPKGGMKTKK 1337
QY 793 -----SPRSP----- 797
Db 1338 SKDYKOPRSPSFGPGPSVLTPRGRPSVSSSEDELTEKPIVIGVPRRDFNDYELIPKQ 1397
QY 798 -----OKKTHYFIAAVER 811
Db 1398 DQEADFGLLDHPBEYVEYKDPYKTDVQALDATSOHLKMGAGDKNTTYFISBEE 1457
QY 812 LWDYGMSSPHVLNRNARQSGSVPOKVKVFOEFTDGTQPLYRGELNEHLGLLGPVIRA 871
Db 1458 ENDYA-CYGQRRLDKTAQNERPTVFRKVVFRYLDSTFSDIRGEMDEHLGLLGPLIKA 1516
QY 872 EVDNIMVTPRNOASRPSYSSLIY-----EEDQAGAPRKNFKVKNETKTYFW 923
Db 1517 EVDQTVMFVFRNARSPYSLHANGVKYLKQMEGLSYDDSPYWKQDDAVPPNGTFTYMW 1576
QY 924 KVQHMAPTKDEPCKAWAYESVDVLEKDVHSGILGLLCHNTNLTN--PAHGRQVTVQE 981
Db 1577 TIIPKSGQNNESDCRTWTIYSAVNPEDINSGLIGLPLLCRGTLDKCPEDR-----E 1631
QY 982 FALFFTFIDETKSYFTENMERNCRAPCNQMEDPTFKENYRFAINGYIMDTPLGLVWA 1041
Db 1632 FULLFTFDENKSWLYEENRQIERKNRRVVM-DPNEQDNLKFPDING-IYSLKGRMY 1689
QY 1042 QDQIRWILUSMGSENIHSTHPSGHVFTVRKKEEYKALYNLYPGVFETVEMLPSPKAGI 1101
Db 1690 TNQLAKWHLINMGSPKDLHSHVHFGQTFINKELKDRHQGVYPLLPGGFATLMLPSKPL 1749
QY 1102 WRVECLLGEHLHAGMSTLF-----LVYSNCKQTPFGMASGHIRDFQI 1143
Db 1750 WQLESREVLGSQQRGMOTLFLFDESYYKRYVCQVLRSLYLIYFAVCDHPLGLISGTVDQEI 1809
QY 1144 TASQYQGWAPKLARLHYSGINAWST--KEPFSWKYDILLAPMLIHGKTKQGRQKPSL 1202
Db 1810 TASDTRQWTPHLARLNTKYNASTSSEPGQLQVDFORPVVISKATQGAQKQFLTHN 1869
QY 1203 YISQFIIMYSLDGKKWQYRGNSTGT-----LMVFF-----GNVDSGKIKHNFNP 1248
```

RESULT 4

```
O43737
ID O43737 PRELIMINARY; PRT; 2224 AA.
AC O43737;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Factor V.
OS Homo sapiens (Human).
OC Karyota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RP Bird C.;
RA Bird C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; Z99572; CAB16748.1; -.
DR HSP; P12259; LC2T.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8 type C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 2.
SQ SEQUENCE 2224 AA; 251673 MW; 1013108D4960EAB CRC64;

Query Match 31.0%; Score 2386.5; DB 4; Length 2224;
Best Local Similarity 26.0%; Pred. No. 2.2e-169;
Matches 592; Conservative 278; Mismatches 483; Indels 923; Gaps 35;

QY 3 RRYLGAVELSDWYMQSDLGELPVDARFPFRPKSPFPNTSVV--YKTLFVFEVTHLPNI 61
Db 32 RQFVAAQGISWRPE-----PINSLSLVSPTSFKIVREVEPY-FKK 75
QY 62 AKRPPPMGLLGTIOAEVYDVTVITLKNWASHPVSHVGVSWKASGEAYDDQTSQR 121
Db 76 EKPQSTISGLLGTPLVAEVDGIIVKHVKADKPLSHIQGIRYKSLSEGAYLDHTFPA 135
QY 122 EKEDDKVFPQGSHTYVYVQVLEKNGPMASDPLCLTYSLVSHVDLVKDLNSGLIGALLVCRE 181
Db 136 EKMDADVAPORETYEWSISEDSGPTDDPPCPTHYIYSHENLIEDFNLSGLIGLLEICK 195
QY 182 GSLAKSKTO-TLHK-FILLFAVEDEGKSWHSEKNSLMQDRDAASARAWPKMHTVNGYVN 239
Db 196 GTLTEGGTQKTFDKQIVLLFAVEDESKSWSSQSS-----LMTVNGYVN 239
QY 240 RSLPGLIGCHRSKVYWHVIGMTPTPEVHSGIFLSEHTFLVRNHRQASLEISPIFLTAQTL 299
Db 240 GTMPDITVCAHDHISWHLGMSGPELFIHFNGQVLEQNHKVSAILTVSATSTTANMT 299
```


[illegible]

Db 261 TQESNRHAIINGVFGNDPELNMCAQKRVAMHLFGMGNEIDVHTAFHGMQLTRGHHT 320
 QY 284 ASLEISPIITFLTAQTLMDLQGLFLCHISSHQHDMAYVYKVDSCPEEPQIRMKNEEA 343
 Db 321 DVANIFPAFTVAEVPWEPGTWLSICQVNSHFDRGMQALYKVKSCSMAPPV----- 372
 QY 344 EDYDDLLTSEMDDVVRPDDNPSFIQIRSVAKKHPKTVWVYIAAEEEDWDYAPL----- 398
 Db 373 -----DLTGT-----KVRQYFIEAHEIQWDY 401
 QY 399 -----VLAPDRSKYQSLNNGPQIRGRKYKVRFMAYTDETKTREAIQHS--GILGPL 452
 Db 402 TGNLREPG--SISDKFFQKSSRIGGTWYKVEAFQDETFOEKWHEEDRHILGPLV 459
 QY 453 LYGEVGDTLIIIFKQASRPYNIYPHGTTDVRPLYSRRLPKGVKHLKDPFPLPGIFK-- 510
 Db 460 IRAEVGDTTQVVFYNRASQPFMSQPHGV-----FYEKDYEGTV--YNDGSSYFGLVAKPF 512
 QY 511 -----YKWTVTVEDGPTKSPRCITRYYSFVNMERDLASGLIGPLII CYKESVDQGNQI 566
 Db 513 EKVYRWTPVPHAGPTAQDPACLTWYFSAADPIRDTNSGLVGPLVCRAGALGADGKQK 572
 QY 567 MSDXRNVILSFEDENRSLTENIQRFLPNPAGVQL-----EDPE-FQASNMHSG 619
 Db 573 GVDKEFFLFTVLDEKNSYSN-----ANQAAAMLDFRLSIEDIEGFQDSNRHAI 625
 QY 620 YPDSI-QLSVCLHEVAYWYIISGAQDTFLSVFSGYTFKWKVYEDTLIFPSSGETV 678
 Db 626 FLFSNLPRLDMCKGDTVAHLLGLGTDTDVHGMFQGNVTQVQGMKGAAMLFPHTFMA 685
 QY 679 FMSMENPGLWILGCHNSDFRNGMTALLKVSSCDKNTGTYEDSYEDISAYLLSKNNAIE 738
 Db 686 IMQPDNLGTFEYCOAGSHREAGRAIYVNSQC----- 718
 QY 739 PRFSQNPVPLKRHQREITRTTLQSDQEBIDYDITISVEMKXEDFDIYDEDNQSPRSFQ 798
 Db 719 -----PGHQ-----ATPRQY 729
 QY 799 KTRHYPIAAVERLDWYGMSS--PHVLNRPAQSGSV-----PQKKVVFQEF 844
 Db 730 QAARIYIMAEVENDYCPDRWERHNSQSKSYGIFLNSKDLGLSGRYKKAVFREY 789
 QY 845 TQGSFTQPLRYGELNHLGILGPIYIRAEVEDNIMVTFRNQAQSRPYSFYS-SLISYEEDQR 903
 Db 790 TDCFTFIPRPTGPEEHLGILGLIKEVGVDILTVVFKNNSRPYSVHAHGVLESTTWP 849
 QY 904 QGAEPKRFVKNETTYFKVQHMAPTYKDEDFCKAMAYFSDVDLEKDVHSGILGPLLV 963
 Db 850 LAEE-----PGEVVTYQWNPERSGPGFNDACSVCWITYYSAVDPIKDMYSGLVGLAI 902
 QY 964 CHTNTLNPAGROVTVQEFALPTTIFDETQSWYFTENM-ERNCRAPCNIMQEDPTFKENY 1022
 Db 903 CQXGILPBGHSDMDREFFALLFIIDENKSWYLENVATHSGQDPGSINLQDETFLSN 962
 QY 1023 RPHAINGYIMDTPLGLVMAQDQIRWYLLSMGNSNENIHSIFHSGHVTVVRKCEYKMA 1082
 Db 963 KMEHANGKLVANLGLTWQGERVAMVYMLANGQVDLHTIHFHABSFLYRNGENYADV 1022
 QY 1083 NLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYVKNKCTPLGMASCHIRDFQ 1142
 Db 1023 DLEPGTFEVEMVNASPGTWLMHCHVTDVHAGMETLFTVFSR-----TEHLSPLT 1073
 QY 1143 ITASQGVQWAPKARLHYSGSINAWSTKPEFSWIKVDLLAPMIHGIKTQAGKQFSSL 1202
 Db 1074 VITKEKAVPPDIE---EGNVKMLGMQPIK--NVEMLASVIV-----AI 1115
 QY 1203 YISQFIMSLDGKQWTVR 1222
 Db 1116 SVTLVLLVLAAGGVWYQHR 1135

RESULT 10

075180
 ID O75180 PRELIMINARY; PRT; 1104 AA.
 AC O75180;
 DT 01-NOV-1998 (T-EMBLrel. 08, Created)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE KIAA0698 protein (Fragment).
 GN KIAA0698;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98403880; PubMed=9734811;
 RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. X.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro."
 RL DNA Res. 5:169-176 (1998).
 DR EMBL; AB014598; BAA31673.2; -.
 DR HSSP; P00450; ICKW.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR002355; MultiCu oxidase2.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
 DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
 FT NON TER 1
 SQ SEQUENCE 1104 AA; 124257 MW; 617C753F766AA152 CRC64;
 Query Match 22.2%; Score 1710; DB 4; Length 1104;
 Best Local Similarity 32.3%; Pred. No. 4.3e-119;
 Matches 397; Conservative 183; Mismatches 432; Indels 216; Gaps 25;
 QY 46 YKTLFVEFTVHLNTAKPRPPMGLGPTIOAEVYDVTVTILKNMASHPVSLHAGVSY 105
 Db 19 YKTIYKEYKDDSVTDEVAQFAMLGFLGVLQAEVGVILLHLKFNATPYTHPHGVY 78
 QY 106 WKASEGAYDDQTSQREKEDDYPPGSGHYTVQVLEKENGPMASDPLCLITYSLSHVDLV 165
 Db 79 EKDSGLYPDGSGPLKADDSVPFGSHTYNTTPEGHAPTADADPACTWIYSHVDAP 138
 QY 166 KDLNSGILGALLVCREGLS---AKETQTLHKFILLFAVFDGKSWH-----SETKNS 215
 Db 139 RDIATGLIGLITCKRGLDGNPPQDQVDHDFLLFSVDENLSHLNENIATYCSDP 198
 QY 216 IMQDRDAASARAPKMTVNGYVNRSLPGLIGCHRSVYVHWI GMGTTPEVHSIFLEGHT 275
 Db 199 ASVDKEDETFQESNRHAIINGFVGNLPELNMCAQKRVAMHLFGMGNEIDVHTAFHGM 258
 QY 276 FLVNRHQASLEISPIITFLTAQTLMDLQGLFLCHISSHQHDMAYVYKVDSCPEEPQL 335
 Db 259 LTRGHITDVANIFPAFTVAEVPWEPGTWLSICQVNSHFDRGMQALYKVKSCSMAPPV 318
 QY 336 RMKNNEAEEDYDDLTSEMDDVVRPDDNPSFIQIRSVAKKHPKTVWVYIAAEEEDWDY 395
 Db 319 -----DLTGT-----KVRQYFIEAHEIQWDY 339
 QY 396 APL-----VLAPDRSKYQSLNNGPQIRGRKYKVRFMAYTDETKTREAIQHS 446
 Db 340 GPMGHGDSGTGNLREPG--SISDKFFQKSSRIGGTWYKVEAFQDETFOEKWHEEDRH 397
 QY 447 --GILGPLLYGEVGDTLIIIFKQASRPYNIYPHGTTDVRPLYSRRLPKGVKHLKDPFPL 504
 Db 398 HLGLIGLGVIRAEVGDITQVVFYNRASQPFMSQPHGV-----FYEKDYEGTV--YNDGSSY 450
 QY 505 PGEIFK-----YKWTVTVEDGPTKSPRCITRYYSFVNMERDLASGLIGPLII CYKES 558
 Db 451 PGLVAKPFVKTVRWTVPVPHAGPTAQDPACLTWYFSAADPIRDTNSGLVGPLVCRAGA 510
 QY 559 VDQGNQIMSDKRNVLFSVFDENRSLTENIQRFLPNPAGVQL-----EDPE-FQAS 611


```

511 LGADGKQKGVDEKFFLLFTVLDEKNSWYSN-----ANQAAALDFFLLSEIDIEGFQDS 563
QY 612 NIWHSINGVYFDSL-QLSVCLEHVAWYILSICAQTDFLSVFFSGYTFKHKVYEDTLTL 670
Db 564 NEMHAINGLFLENPLDCKDGTVAWHLLGLGTETDVHGVVFMFGQNTVQLQMRGAAML 623
QY 671 PFPSGSTVPMSENPGLMILGCHNSDFRNRGHTALLKVVSCDKNTGDYDESDYEDISAYL 730
Db 624 PHTFVMAIQPDNLGTFPIYCOAGSHRAGMRAIYVNSQC----- 664
QY 731 LSKNNAIEPRFSQNPPLVKRHOEITRKTLOSQDEIDYDDTISVEMKEDFDIYDEDE 790
Db 665 -----PGHQ----- 668
QY 791 NOSPRSFQKTRHYFIAAVERLWDYQWSSS-PHVLNRAQSGSV-----POF 836
Db 669 -ATPRQYQAARIYYIMAEVEWDYCDPRSWEKREHNSQEKDSGYIFLSNKGDLIGSRY 727
QY 837 KKVVFQRFYDGGSTQPLRGELNEHLLGPYIRAEVEDNIWYTFNQASRPSVYS-SL 895
Db 728 KKAVFREYDGTFRIPRPTGPBEHIGILGPLKGEVDLITVVFKNASRPYSVHAHGV 787
QY 896 ISYEEDQRCGAEPKRVFKNPTKTYFWKVQHMAPTKDFDCKAWAYFSDVDLEKDVHS 955
Db 788 LBSSTVWPLAEE-----PGEVVYQWNIPEPSRGPNDACSVMYIISAVDPIKDMYS 840
QY 956 GLIGPLLVCHTYTLNPAHQGVTVQSFALFFTFDETKSWYFTEHM-ERNCRAPCNIQME 1014
Db 841 GLVGPLAIQCKGLEPHGGRSDMREFFALLFLFDENKSWYLEENVATHGSDQGSINLIQ 900
QY 1015 DPTFKENYRFHAINGYIMDTPLGLVMAQQRIRWYLLSMGNSNENIHSIHFSGHVFTVRKK 1074
Db 901 DETFLASNKGHALNGKLYANRLGLTYQSERVAVWYMLANGQVDLHTIHFHAEFLYRNG 960
QY 1075 EYKMKALYNLYPGVFTVEMLPSKAGIWRVECLIGELHAGHGSTLFLVYSNKKQTPLGMA 1134
Db 961 ENYRADVDLFPQTPEVFEVWASNPCTWLMHCHVTVDHVHAGMETLTFVFSR----- 1011
QY 1135 SGHIRDFQITASQYQGWAPKLARLHYSGSINAWSTKEPFSWTKVDLLAPMIITHGKTG 1194
Db 1012 THELSPLTWITKETEKAVPFRDIE---EGNVKMLGMQIPIK--NVEMLASVLV----- 1059
QY 1195 ARQKFSLSYISQFIIMYSLDGKKWQTYR 1222
Db 1060 -----AISVTLLLVVVALGCVVWYQHR 1081

RESULT 11
Q9JUL97 PRELIMINARY; PRT; 1084 AA.
AC Q9JUL97;
ID Q9JUL97;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GPI-anchored ceruloplasmin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxId=101116;
RX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Sprague Dawley; TISSUE=Brain;
RC MEDLINE=20127919; PubMed=10660599;
RX Patel B.N., Dunn R.J., David S.;
RT "Alternative RNA Splicing Generates a Glycosylphosphatidylinositol-
RL anchored Form of Ceruloplasmin in Mammalian Brain.";
RL J. Biol. Chem. 275:4305-4310(2000).
DR ENEL; AF202115; AAF34175.1; -.
DR HSPF; P00450; iKCW.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR002355; MultiCu oxidase.
DR Pfam; PF00394; Cu-oxidase; 3.

```

```

Db 849 RYIIQIPERSAGCTEDSPCIWAIYSTVDRVKDYISGLIGLIVCRKSYKVNPK--- 905
QY 976 QYVQEFALFTTFIDETKSYFTENMERNCRAPCNQIOMEDPTFKENYRFAHNGYIMDTL 1035
Db 906 --KMBEYSLFLVFDENESWYLDNINTYSDHPEKVNKNBEFTESNKMHAINGKMFNL 963
QY 1036 PGLVMAQDQIRIWMYLLSGNSNENIHSIHFSGHVFTVRKESYKMAIYNLYGVETVEML 1095
Db 964 QGLTHVHGDEVNMYWAMGNEIDLTHVHFHGHSPQYKRGHSHSDVDLFPFGTYQTLEMF 1023
QY 1096 PSKAGIWRVECLIGELHAGMSTLFLVYSNKKQT 1129
Db 1024 PQTGPGWLLHCHVTDHAGMVTYTVLPNQASS 1057

RESULT 12
Q9XT27
ID Q9XT27 PRELIMINARY; PRT; 1048 AA.
AC Q9XT27;
DT 01-NOV-1999 (TREMELrel. 12, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE Ceruloplasmin.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=99384006; PubMed=10452945;
RA Lockhart P.J., Mercer J.F.B.;
RT "Cloning and expression analysis of the sheep ceruloplasmin cDNA.";
RL Gene 236:251-257(1999).
DR EMBL; AF134814; AAD41477.1; -.
DR HSSP; P00450; 1KCW.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR002355; MultiCu oxidase2.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
SQ SEQUENCE 1048 AA; 119125 MW; 925F16D7B0549CBB CRC64;

Query Match
Best Local Similarity 32.2%; Pred. No. 3.2e-112;
Matches 378; Conservative 177; Mismatches 417; Indels 202; Gaps 25;

QY 3 RRYILGAVLSWDYMQSDIGE---LPVDARPPRPVKSPFPNTSVYKTLFVEFTVILF 59
Db 22 KHYTIGIITAWNY-ASDHAEEKLISVDTESNIYLONGPNRIGSVYKAVLYQYTDENF 80
QY 60 NIAPRRPMMGLIGPTIOAEVDTVITLKNASHPVSLHVGVSYKASGAEVDDOTS 119
Db 81 RTVIEKPWLGFLGFIKAETGDKVTVHLKNFASPRYTFHAGLYYEHGALYPDNTT 140
QY 120 QREKEDKYPPGSSHTYVQVLKNGPMASDPLCLTYSLYSHVDLVKDLNSGLICALYVC 179
Db 141 DLQAKADKYQPGFQCLYILHANPEQPGBEDSNCTRIYHSHIDAPKDIAASGLIGLTHC 200
QY 180 REGSLAKEKTQTLHK-FILLFAVFEKSWH-----SETXNSLMQDDAASARAWP 229
Db 201 KQSLDEEKEKNIDKEFVVMFVVDENUSWYLEENIKYCSPEKVEQDNEDFOESN--- 257
QY 230 KMHTVNGYVNRSLPGLIGCHRSVYVHWVIGMTTPEVHSIFLEGHTFLVRNHRQASLEIS 289
Db 258 RMYSVNGYAFGLSLPGLSMCAEDRVKWLFGMGNEIDVHAFFHGQVLTSKVRVDTINLF 317
QY 290 PITFLTAQTLMDLQGLFLFCHISSHQHDGMAYKVKVDSCEPQOLRMKNNEEADYDD 349
Db 318 PATLFAFVAQNPQGMWLSQNLNHLKAGLQAFVFWQDCKSS----- 361

```

```

QY 350 LTDSEMDVVRFDODNSPFIQIRSVAKGPKTWVHYIAAREEDWDYAPL-----VL 400
Db 362 -----SEDN-----IHGKNVRH-----YIAAEVIMWYAFSGIDAFKKNLR 399
QY 401 APDRSYKQYLLNNGSPQIRGRYKRVFMAYTDTF---KTRERAIQHSGLIPLLYGV 457
Db 400 APGSAS--EAFPEQGPTRIGGSYKXKLVREYTDASFNSQKGRGPEEHLGILGPVIAEV 457
QY 458 GDTLLIIPKQASRPNIYPHGI-----TDVRLPSRRLPKGVKHLKOPPIIPGIF 509
Db 458 GDTIRVTFNKAAHPLUSIEPIGVVRVDKNNEGTYTPTSGPPFSG-SH-----VAPKGT 511
QY 510 KYKWTIVTVDGPTKSDPRCLTRYISFVNMRDLASGLIGLPLLCYKESVDQGNQIMSD 569
Db 512 TYEMTVPEVGYTKDPVCLAKWYS--GSTKDIFGLIGPMKI CNGSLLANGRLKXND 569
QY 570 KRNVIILSVFDENRSWYLTENTQRELPNAPAGVOLBEDPBPQASINIMHSINGYVDSLQ-LS 628
Db 570 KEFYLFTVFDENESLLDDNIMFTTAPDQVDKENEDFOESNKMHSNMNGFYGNQPLS 629
QY 629 VCLHEVAVWYILSIGAQTDFLSVFESGYTFKHKWVYEDTLTPFESGETVFMESMENPGLW 688
Db 630 MCQGSVMWYLSAGNEVDIHGIFSGNTYILSGERRDTANLFPQISLSLFPQDTAGTF 689
QY 689 ILCHNSDFRNRGMTALLKVSXCDKNTGYEDSYEDISAYLLSKNNAIEPRSFQNPV 748
Db 690 DVECLATDHYTGMMQKYTVSQGQRS-----EDLYYL----- 723
QY 749 LKXHQREIRTTLOSQBEIDYDDTISVEMKEDFDIYDEDNQSPRSFQKTRHYFAA 808
Db 724 -----CERTYITAA 732
QY 809 VERLDYNGMS-----SSPHVLRNRAOAGSV-----POFKKWFOEFTDGSFTOPLYR 855
Db 733 VEVEWDYSPSRKWEKELHLQNLNAPLDEKEFYIGSKYKVVYRQFTDSTFQVYER 792
QY 856 GEINELHGLIGPYIRAEVDNIMVTRNQAASRPYFYSLSIYEBEDQOGASPRKNFKVP 915
Db 793 KGEELHGLIGLPOLHADVGDKVNIIPKKNATPRYSIHAHVKTESST---VTP---TAP 845
QY 916 NETKTYFMKQVHMAPTKDEEDCKAWAYPSDVLKDVHSGGLIGLPLVCHTWTL---NPA 972
Db 846 GEIRTYINKLIPERSGAGMDSPCIWYVYISTVDRVKDLFSGGLIGLIVCRKHYLVSNP- 904
QY 973 HGRQVTVQEFALFTTFIDETKSYFTENMERNCRAPCNQIOMEDPTFKENYRFAHNGYIM 1032
Db 905 ----IKKLEFSLFLVFDENESWYLDNINTYSDHPEKVDKANEEFMESNKMHAINGRMP 960
QY 1033 DTLPLVMAQDQIRIWMYLLSGNSNENIHSIHFSGHVFTVRKESYKMAIYNLYGVETV 1092
Db 961 GNLQGLTM-----HVGNEVDLSVHFHGHSPQYKRGHSHSDVDLFPFGTYQT 1009
QY 1093 EMLPSKAGIWRVECLIGELHAGMSTLFLVYSNKK 1126
Db 1010 EMTPTGPIWLLHCHVTDHAGMVTYTVLPNE 1043

RESULT 13
Q8BQ43
ID Q8BQ43 PRELIMINARY; PRT; 355 AA.
AC Q8BQ43;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Coagulation factor VII (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Boreal root ganglion;
RX MEDLINE=22354683; PubMed=12466851;

```


